



SCMS SCHOOL OF ENGINEERING & TECHNOLOGY

PUBLICATION DETAILS 2020

SI No:	Name	First Author	Second Author	Third Author	Fourth Author	INDEXING
1	Dr.Sunil Jacob	PEC2001,PEC2002,PEC2003	PCSE2001,PCSE2003,		PCSE2004	SCI
2	Dr.Saira Joseph		PEC2003	PEC2002		SCI
3	Vinoj PG				PEC2002	SCI
4		PCSE2003,PCSE2004,PCSE2005(6), PCSE2006(5)	PEC2002,PEC2003	PEC2001	PCSE2001	SCI
5	Dr. Varun Menon	,PCSE2008(6),PCSE2011(6), PCSE2013(5),PCSE2014(5), PCSE2015(5)	PCSE2020(SCO)	PCSE2007, PCSE2012, PCSE2017 PC2018	PCSE2009,PCSE2010, PCSE2016	SCI
6	Divyanathn K	PEE2001				SCOPUS
7	Divya M S	PBSH2001,PBSH2003(UGC), PBSH2004,PBSH2005				SCOPUS
8	Febini Joseph	PBSH2002				UGC CARE
9	Rathesh Menon		PCE2001(SCO)	PCE2004		SCI
10	AKhila M	PCE2002(SCI),PCE2005(SCI)	PCE2003(SCO)			SCI
11	Dr. Vinod P				PCSE2002	SCI
12	Nitty Rose Augustine			PBSH2001		SCOPUS

Total Publication for the calender year 2020	33
---	-----------




DR. PRAVEENDAL C.J.
 PRINCIPAL
 SCMS SCHOOL OF ENGINEERING & TECHNOLOGY

Article preview

Abstract

Introduction

Section snippets

References (52)

Cited by (25)



Process Safety and Environmental Protection

Volume 138, June 2020, Pages 337-348



Failure mode effect and criticality analysis using dempster shafer theory and its comparison with fuzzy failure mode effect and criticality analysis: A case study applied to LNG storage facility

Manoj Jose Kalathil ^a ✉, V.R. Renjith ^a, Nitty Rose Augustine ^b

Show more ▾

+ Add to Mendeley 🔗 Share 🗨️ Cite

<https://doi.org/10.1016/j.psep.2020.03.042>

[Get rights and content](#)

Abstract

Author

X

Nitty Rose Augustine

[View in Scopus](#)

SCMS, School of Engineering and Technology, Kerala, India

More documents by Nitty Rose Augustine

Provided by Scopus

[Failure mode effect and criticality analysis using dempster shafer...](#)

Process Safety and Environmental Protect...
Kalathil, M.J., ..., Augustine, N.R.

Activate Windows

Go to Settings to activate Windows.

FEEDBACK



RESEARCH ARTICLE

Vol. 7. Issue.1. 2020 (Jan-Mar)

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA
2395-2628(Print):2349-9451(online)

A STUDY ON WOMEN EDUCATION: EDUCATE AND EMPOWER

DIVYA MS

Assistant Professor, English Department, SCMS School of Engineering and Technology, Ernakulam

Email: divya718@gmail.com



Article information

Received:27/02/2020

Accepted: 15/03/2020

Published online: 21/03/2020

doi: [10.33329/ijelr.7.1.179](https://doi.org/10.33329/ijelr.7.1.179)

ABSTRACT

Education is the key to women's empowerment. When education gets denied there isn't much to happen in her life. As usual she gets married and have children at a young age, work in unpaid or low-paying jobs, and rely on economic support from their husband or family. Without education, their future and their family's future gets limited. According to the Malala Fund, there are over 130 million girls worldwide who are not in school. Has anyone thought about them or their future? There might be an unpolished gem among them. Studies have found that if every girl completes 12 years of education, child marriage would drop by 64% and health complications from early pregnancy, like early births and child deaths, would drop by 59% and 49%, respectively. Educating women and girls also improve the economic status of a nation, which will reduce the risk of war and terrorism and can be a better harbinger of future. There are still many barriers preventing girls and women to pursue and complete their education like fees, distance or lack of transportation, being forced to work and provide for their families, being forced to marry and have children, or conflict in their hometown or country. The most difficult obstacle is the mentality of the family towards a girl child. They are denied education only because she is born as a girl. The United Nations found that as girls reach secondary school, their enrolment rates decline significantly. Only 39% of the countries have equal proportions of boys and girls enrolled in secondary education. In developing countries, 35% to 85% of girls are forced to stay at home from school to take care of their younger siblings and to manage the house while their brothers are provided with higher education. To reach the competition level and to expand their professional opportunities, women need the same experiences and skills, making post-secondary education an essential part of women's empowerment. Higher education instils them with the knowledge, competence and experience that are necessary to get involved in government, business, or even in a civil society. With higher education, women and girls have better access to health information and other beneficial services which in turn will only help the family or generation to grow and develop. We need more and more sheers to be the torch bearers of future.

Key words: Empowerment, Post-secondary education, Shero, Professional opportunities

“You educate a man; You educate a man.

You educate a woman; You educate a generation”-Brigham Young

We are living in a highly competitive world where survival is the only mantra to be focused. Such a fast growing world hardly differentiates between men and women. Education is the only key factor that brought towards such a change. Even when women have access to education, there can be many other factors that can make it difficult for her to take full advantage of those opportunities. It's women who still carries the cultural burden of being the primary homemakers and caregivers. This unpaid “second shift” means that they have less time and energy to dedicate to their studies. Family responsibilities doesn't become burden for them but that is the prime factor that is delimiting them from their higher education or carrier growth. When women become the sole providers for their families, they need to fulfil it with many other factors like being victims of domestic violence, threats like financial and from profession which becomes much more difficult for them to handle. The immense pressure that she is handling becomes worthless when it is labelled under the synonym ‘mother’ or ‘wife’. Today's era has witnessed a lot of change when it comes to single girl child nuclear family. Deciding voices are only that of their parents. Thus providing wings for them to fly high and capture their aim. Their dreams can't be in fetters and who knows there might be an unpolished gem that is being soiled. Women in India have been treated with utmost respect and dignity since time immemorial. For a nation to advance, empowering women is crucial. India was blessed with visionary women who broke the fetters of gender norms in every sector. Its long back we heard about the denial of education to women. A girl child was considered curse than a boon. Never to live in a world where mothers and sisters are dragged out of their homes and raped. Wherein she becomes both the victim and accused and its said by “them”, it's all because of her “out of the box” attitude that caused her this. Is this the caring that we need to give to our sisters, rather than rending a helping hand? We can't even skim through the news heads that is going on these days. What if such an accusation comes out, it needs to face a frequent trail from the social medias and channels who will make them even worse? Is the only solution, to remain silent and bear the tragic effects or to have a prolonged grudge and to end up in depression throughout the life? It's an open ended question towards the secularist nation. Here I would like to quote the famous wording of Michelle Obama,

“When girls are educated, their countries become stronger and more prosperous”

Let the nation realize her power of golden touch. She is to succeed where ever she goes, don't curtail her from her doing what is right. Listen to her inner voice that itself will keep the nation going. We do respect our mothers and all woman is an incarnation of her, our mother Earth. Its only she who can protect, provide, love, sacrifice selflessly. Split of the moment can we see her incarnations of a loving Sita, ready to sacrifice everything for her love and family, Durga the fearest of all who battle against evil for good. For us this is our mother, both deities in one, who loves and cares us with one hand and get furious with another. Still we love her and keep no grudges. Then how can we ever think of harming her. Where has our brains and minds gone, ready to stab the same womb from where we came? Is this what we call progression? Are we the real citizens bearing torches for a future developing nation? How can we sustain by spiting on our own veins? When will all these quires be dissolved? The only solution for all this is only education, it opens the eyes to a new world of realities and hopes for a better nation. It's clearly evident that the only means to attain empowerment is through education. Change should begin from her and its only she who can bring changes. Here I would like to cite the story of Rani Padmini a legendary 13th – 14th century Queen (Rani) of the Mewar kingdom of present day India. She was the wife of King Ratan Sen, captured and imprisoned by Delhi's sultan Alauddin Khalji. Alauddin Khalji became enamoured with Padmavati's beauty and decided to siege Chittor to obtain Padmavati. Before Chittor was captured they had to face defeat against Khalji, she and her companions committed Jauhar (self-immolation) thereby defeating Khalji's aim and protecting their honour. Coupled to the *Jauhar*, the Rajput men died fighting on the battlefield. Throughout the novel *Padmini: The Spirited Queen of Chittoor*, Mridula Behari tried to explain the immense beauty and knowledge that Padmavati possessed. It's only because of her education and knowledge, she had the courage to face such a situation where her husband the Rana itself was helpless. She was bold enough to convince everyone and to equip them for a war against the evils. There are many instances in the novel where we find her indulged in reading the ancient scriptures and getting mesmerised by that. It's the education that

brought her into a new world where she understands that action speaks more than words. There is always an urge for action than mere talks, that won't lead you anywhere. Education delimits the boundaries of caste, creed, gender, finance and what not. How true it is to quote Malala Yousafzai's wordings here,

"If one man can destroy everything, why can't one girl change it?"

As Prime Minister Narendra Modi said on the launch of the expanded 'Beti Bachao Beti Padhao' (March 8, 2018): "Daughters are not a burden but the pride of the whole family. We realise the power of our daughters when we see a woman fighter pilot. The country feels proud whenever our daughters bag gold medals, or for that matter any medal, in the Olympics." This is the time only when we think oh! are they capable of bagging a gold and that too for the nation? When they come to lime light we find applauds and cheering from all over the nation. But have ever we imagined the trials or pains that they have gone through to achieve that? Same is the case with the other gender, but they are always on safe side of the scrutiny eyes of the society. It's a boy, no matter if he is alone or late to home or traveling late or alone. The world is always open to him than to her. We need more Sheros to be an inspiration and to motivate others to come out from their shells that is encircling them towards darkness. Engaged in a combat to the march for equality with our sisters and mothers, let us understand the theme of Women's Day: "An Equal world is an enabled world: realizing women's power." After the adoption of the Beijing Agenda for Action, UN has set the year 2020 as a key year for assessing international progress towards achieving gender equality and human rights for all women and girls. The Ministry of Human Resource Development (MHRD) has much triggered up their progression under the leadership of PM Modi in providing equal opportunities. It is with immense happiness we can say that due to the Swachh Bharat Mission, 14,67,679 schools now have a functioning girl's toilet, an increase of 4.17 percentage points in comparison to 2013-14. The impact of the mission has resulted in an increase in enrolment of girls by 25 percentage points in 2018-19 from 2013-14. These figures get dimmed in a society which is indulging much more in a political game for their sustenance. We living in the safe shelters haven't ever thought about our sisters who is being victimized and marginalized. Minister of Finance Nirmala Sitharaman applauded the performance of Beti Bachao Beti Padhao in her speech on the budget: "Gross enrolment ratio of girls across all levels of education is now higher than boys. At the elementary level it is 94.32 per cent as against 89.28 per cent for boys, at the secondary level it is 81.32 per cent as compared to 78 per cent and at the higher secondary level girls have achieved a level of 59.7 per cent compared to only 57.54 per cent." The MHRD has approved 5,930 Kasturba Gandhi Balika Vidyalayas, which are girls' residential schools and have an enrolment of 6,18 lakh students, to increase equality of access and opportunity for girls. The National Incentive Scheme for Girls for Secondary Education has approved an incentive sum of Rs 8,56 crore for the 28,547 beneficiaries. According to the scheme Rs 3,000 is being deposited under the age of 16 in the name of deserving unmarried girls and entitles them to withdraw it along with interest in reaching 18 years of age and passing Class X. Besides an improvement in the girl's gross enrolment rate in schools, the educational outcomes and accomplishments have also improved.

Let's go ahead with this initiative for our sisters than to get involved in the cheap political drama that is being happening in our nation. There is considerable evidence that women's education and literacy tend to reduce the mortality rates of children. It's indeed true what Malala Yousafzai pointed out:

"We realize the importance of our voice only when we are silenced"

In accordance with the celebration of India's success in improving gender equality in the education system, much greater and collective efforts are needed to achieve the Sustainable Development Goal of eliminating gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations. History of Indian women is full of pioneers, who broke gender barriers and worked hard for their rights and made advances in politics, arts, science, law, etc. Let us cite few examples of our pioneers who made us think beyond Anandibai Gopalrao Joshi, who became the first Indian female physician in the year 1887. She was also the first Indian woman who get training in Western medicine and the first woman to travel to the United States of America. Arunima Sinha, is the first female amputee to climb Mount Everest. She is also the first Indian amputee to climb the Everest. She was a national level volleyball player who in 2011 was pushed by

robbers from a running train as she defied them. After meeting this accident, one of her legs was amputated below the knee. Arati Saha became the first Indian and Asian woman to swim across English Channel in the year 1959. She also became the first female sportsperson awarded with Padma Shri in 1960. Mother Teresa founded many Missionaries of Charity, a Roman Catholic religious congregation, giving her life to social work. Indira Gandhi became the first woman Prime Minister of India and served from 1966 to 1977. Indira Gandhi renowned as the "Woman of the Millennium" in a poll which was organised by BBC in 1999. In 1971, she became the first woman to receive the Bharat Ratna award. Justice M. Fathima Beevi became the first female judge to be appointed in the Supreme Court of India in the year 1989. In her autobiography she had said about the immense suffering that she had faced to reach such a highest peak. Her father was the only person who supported her thorough out her journey. Kalpana Chawla, the first Indian woman who reached in space. As a mission specialist and a primary robotic arm operator, she went into space in 1997.

We can move to our present Sheros starting with Mithali Raj, the first woman to score a double hundred in Test Cricket against New Zealand at Wellington, 2004. She was the first to achieve this landmark in the world. Pratibha Patil became the first woman President of India and held office from July 2007 to July 2012. Kiran Bedi, joining Indian Police Service (IPS) in 1972, she became the first woman officer in India. Moreover, later in 2003, Kiran Bedi also became the first woman to be appointed as the United Nations Civil Police adviser. Anjali Gupta is the first female flying officer in the Indian Air Force to be court martialled. She used to work for the Aircraft Systems and Testing Establishment unit in Bangalore. Anjali completed her Masters of Philosophy in Sociology from the Delhi University and was first posted at Belgaum in 2001. Sania Mirza, a professional tennis player, became the first ever Indian woman to win the Women's Tennis Association (WTA) title in 2005. Later in 2015, Sania Mirza became the first Indian woman titled as rank number one in WTA's double rankings. Saina Nehwal became the first Indian women to win a medal in Badminton at 2012 Olympic Games. Later in 2015, she became the first Indian woman to secure no. 1 position in world rankings. Mary Kom, is the only woman boxer who has won a medal in each of the six World Championships. She was the only Indian woman boxer who qualified for the 2012 Olympics and became the first Indian woman boxer to win a gold medal in Asian Games in 2014. Cited just a few but more hands of achievements are behind which are yet to be recognised and appreciated. They have set a model for us to think and act beyond.

Our government has also taken much initiative for protecting the rights of education for girls. India Post or Department of Posts, the postal system of the country, offers several savings schemes with different interest rates. The **Sukanya Samriddhi Yojana**, one such savings scheme offered by India Post, is a deposit scheme for the girl child that can be opened in any of the leading banks and post offices across the country. In such schemes the child is getting the benefit for future education and for her marriage or future life. There is a platform called WE, that is an empowerment program through which woman are trained to form self-organized and self-managed savings groups, each consisting of 15-25 members. Their aim is to develop individual empowerment and increase their access to financial resources which is the prime element for eradicating poverty. All the members meet weekly to make decisions and interact in life-skills training, discuss various issues of mutual interest. They not only give a platform for awareness but also make an effort to join together and take action to improve their lives and communities. We too need to make more and more efforts than *ME TOO* to raise our voice against the inhuman oppression and injustices that is happening worldwide. Amendments to laws are must as Judiciary is the ultimate power which we believe and rely on. For a common man judiciary is the only hope or last and final resort. We all need to respect our law than fearing it. If the administrators of law are more channelized and less corrupted, we don't have to wait for justice. It's absolutely true to say the famous phrase "Justice delayed is justice denied". Let us be the harbingers of a brighter future that initiates the slogan justice and tranquility. Let the coming era witness a world devoid of corruption, discrimination, poverty, illiteracy.

"Empower yourselves with a good education, then get out there and use that education to build a country worthy of your boundless promise"

-Michelle Obama

Bibliography

“Women empowerment”. Indian Express article. <https://indianexpress.com/article/opinion/columns/beti-bachao-beti-padhao/article/6297784>.

“First Indian Women”. India Today. <https://www.indiatoday.in/education-today/gk-current-affairs/story/the-first-indian-women-312243-2016-03-08>

“One Health and Disease: Tick-Borne.” *National Park Service*, U.S. Department of the interior, <https://www.nps.gov/articles/one-health-disease-ticks-borne.htm>.

Behari, Mridula . <https://www.amazon.in/Padmini-Spirited-Chittor-Mridula-Behari-ebook/dp/B0774NH95V>

Behari, Mridula. *Padmini: The Spirited Queen of Chittor*: Penguin publishers, 2017.

Numerical Study on Cyclic Loading Effects on the Undrained Response of Silty Sand



M. Akhila , P. C. Jithesh, K. Rangaswamy, and N. Sankar

Abstract The soil liquefaction is a major earthquake disaster which causes tremendous damages to all infrastructure facilities. The examples during past earthquakes have shown the evidence of liquefaction-induced ground failures in fine-grained soils. Until recently, liquefaction-related studies concentrated on clean sands believing that only sands are susceptible to liquefaction. However, the earthquakes like 1976 Tangshan earthquake, the 1989 Loma Prieta earthquake, the 1999 Kocaeli earthquake, the 2010 Chile earthquake, and the 2011 Christchurch earthquake, etc., showed that sand with fines could also liquefy. The present study deals with the numerical simulations on cycling loading effects on the undrained response of silty sand. The material parameters for the numerical model are found after conducting basic experimental tests. The response of silt sand under the undrained condition of cyclic triaxial loading is analyzed using the hypoplastic constitutive model. The influence of soil parameters, i.e., void ratio/relative density and consolidation pressure level on undrained response of soil is examined from model simulations.

Keywords Hypoplastic model · Silty sand · Undrained response

1 Introduction

The hypoplastic constitutive model has been utilized for all geotechnical applications of field studies since 1985. The original version of the hypoplastic model is coined by Kolymbas (1985) and is improved in further versions. The present improved version of the hypoplastic model is more advanced over the elasto-plastic models. P.-A. von Wolffersdorff (1996) has described the mathematical formulations involved in

M. Akhila

Department of Civil Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala, India
e-mail: akhila144@gmail.com

P. C. Jithesh (✉) · K. Rangaswamy · N. Sankar

Department of Civil Engineering, NIT Calicut, Kozhikode 673601, Kerala, India
e-mail: pcjithesh08@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

M. Latha Gali and R. R. P. (eds.), *Geotechnical Characterization and Modelling*, *Lecture Notes in Civil Engineering* 85, https://doi.org/10.1007/978-981-15-6086-6_86

1067

1 **PM₁₀ source identification using the trajectory based potential source** 2 **apportionment (TraPSA) toolkit at Kochi, India**

3 Afifa K. Shanavas¹, Chuanlong Zhou², **Ratish Menon¹**, Philip K. Hopke^{2,3*}

4

5 ¹Dept. of Civil Engineering, **SCMS School of Engineering and Technology, Karukutty,**
6 Ernakulam, Kerala, India - 683582.

7 ²Center for Air Resources Engineering & Science, Clarkson University, Potsdam,
8 NY 13699 USA.

9 ³Department of Public Health Sciences, University of Rochester School of Medicine and
10 Dentistry, Rochester, NY 13642 USA

11

12 **Abstract**

13 A recently developed open source tool kit named the Trajectory based Potential Source
14 Apportionment (TraPSA) was used to identify the sources of respirable particulates at Kochi,
15 India. Using 24-hour average particulate matter data from samples collected at five regulatory
16 monitoring stations over the five-year period from January 2011 to October 2016, local and
17 regional scale analyses were made. Concentration field analysis was performed using back
18 trajectories generated by Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT)
19 model with inputs from atmospheric reanalysis data. Conditional bivariate probability function
20 analyses were made using local meteorology data to identify the influence of local sources. Most
21 of the stations indicated the contribution from local traffic activities during low wind conditions
22 and from a nearby industrial area especially during high speed winds. Back trajectory analysis
23 identified potential source areas in Kerala as well as in nearby state of Tamil Nadu as contributing
24 to the air quality at Kochi. Arabian sea on the western side was also observed to be a potential
25 source area for Kochi. The study demonstrated the utility of TrapSA as a tool for deriving
26 information about the potential source areas affected particulate matter mass concentrations.

27

28 **Keywords:** PM₁₀, Air pollution, TraPSA, Back trajectory receptor model, HYSPLIT, Kochi

*Corresponding Author Email: phopke@clarkson.edu

Cite this article

Ranga Swamy K, Akhila M and Sankar N
Effects of fines content and plasticity on liquefaction resistance of sands.
Proceedings of the Institution of Civil Engineers – Geotechnical Engineering,
<https://doi.org/10.1680/jgeen.19.00270>

Research Article

Paper 1900270
Received 17/11/2019;
Accepted 24/06/2020

Keywords: dynamics/geotechnical
engineering/seismic engineering

ICE Publishing: All rights reserved

Effects of fines content and plasticity on liquefaction resistance of sands

K. Ranga Swamy PhD

Associate Professor, Department of Civil Engineering, NIT Calicut, Calicut, Kerala, India

M. Akhila PhD

Assistant Professor, SCMS School of Engineering and Technology, Ernakulam, Kerala, India (corresponding author: akhila144@gmail.com)
(Orcid:0000-0002-0514-0841)

N. Sankar PhD

Professor, Department of Civil Engineering, NIT Calicut, Calicut, Kerala, India

A detailed experimental programme was conducted to evaluate the liquefaction susceptibility of non-plastic and low-plasticity soils subjected to cyclic loading under undrained triaxial loading conditions. The study mainly focused on examining the influence of the amount of fines and plasticity indices on the liquefaction resistance of sands. After mixing silt and clay fractions into fine sand, 16 soil combinations were prepared. The silty sands contained up to 40% non-plastic fines and the low-plasticity soils contained a clay fraction of 5–40%. Each cylindrical soil specimen was constituted to a medium relative density and saturated specimens were subjected to a confinement pressure of 100 kPa. The consolidated specimens were then subjected to various levels of cyclic stress amplitudes using a sinusoidal wave load form at a frequency of 1 Hz. The results showed that both the non-plastic and low-plasticity clay soils were less resistant to liquefaction than the fine sand. The soils belonging to categories SM and SC (silty sand and clayey sand, respectively, as per Indian standard soil classifications) were susceptible to liquefaction if the fines passing through a 75 μm sieve were $\leq 40\%$, the liquid limit was $\leq 40\%$, the plasticity index was < 15 and the saturated water content was about 0.86 times the LL.

Notation

C_c	coefficient of curvature
C_u	uniformity coefficient
D_{50}	mean size
e_c	consolidation void ratio
e_o	initial void ratio
e_{\max}	maximum void ratio
e_{\min}	minimum void ratio
e_{sk}	sand skeleton void ratio
N_L	number of cycles to liquefaction
G	specific gravity
w	water content
Δu	change in pore water pressure
ε_A	axial strain
σ_3	effective applied consolidation pressure

1. Introduction

Liquefaction-induced soil failures in earthquakes have been known to occur in several soil deposits in a loose to medium-density state containing non-plastic to low-plasticity fines. In general, those types of soils cause a rapid build-up of excess pore pressures during seismic excitations. It is difficult to dissipate the excess pore pressures within a short duration of an earthquake event due to the presence of small voids in fine-grained soils. Therefore, foundation soils containing non-plastic and low-plasticity clay fractions underneath buildings and geotechnical structures are susceptible to liquefaction during an earthquake or other dynamic event. Liquefaction causes severe damage to building structures, the soil and

soil-retaining structures, in the form of settlements, lateral spreading, tilting, ground damage and so on. The prevention of such disastrous damage is thus required and current research has focused on understanding the mechanisms and finding suitable mitigation measures.

A review of the literature on the undrained response and liquefaction susceptibility of non-plastic and plastic soil mixtures reveals that unique conclusions have not been found. The liquefaction resistance of sand may vary with the fines content (FC) and the plasticity of the fines. The effects of the non-plastic FC on the liquefaction strength of sands have been extensively investigated, with contradictory findings on the basis of comparisons of liquefaction resistance such as the (global or total) void ratio, skeleton void ratio, relative density and so on. Based on in situ tests, Seed *et al.* (1985) reported that the presence of fines induces an increase in liquefaction resistance. However, with an increase in the FC, some laboratory investigations found an increase in the liquefaction resistance (Amini and Qi, 2000; Chang *et al.*, 1982; Dezfulian, 1984; Fei, 1991; Vaid, 1994) while reversal behaviour was observed in other studies (Finn *et al.*, 1994; Kuerbis *et al.*, 1988; Lade and Yamamuro, 1997; Zlatovic and Ishihara, 1997). The liquefaction resistance of sands has been reported to decrease up to a certain FC but then increases with a further increase in FC (Koester, 1994; Polito and Martin, 2001; Troncoso, 1990). According to Shen *et al.* (1977) and Kuerbis *et al.* (1988), the sand skeleton void ratio is the best parameter to evaluate the liquefaction resistance of non-plastic soils. Based on a review of various studies, Carraro *et al.*

Numerical Study on the Undrained Response of Silty Sands Under Static Triaxial Loading



M. Akhila, K. Rangaswamy and N. Sankar

Abstract The silty soils are more susceptible to liquefaction, even under static loading, than the coarse sands. Pore pressure developed during dynamic events may not dissipate easily due to the presence of more number of small voids. Hence, the rate of pore pressure build-up under static/dynamic loading conditions is much faster in silty sands, which lead to a reduction in the soil strength. This phenomenon may be assessed in terms of either contraction or dilation behaviour under triaxial loading. Therefore, it is necessary to analyse the undrained response of silty sands under triaxial loading so that the damages occurring during future dynamic events may be predicted. The present study involves both the experimental and numerical simulations on various silty sands, which contain 0, 10, 20, 30 and 40% silt fines. Initially, experimental static triaxial testing was performed to determine the undrained response of silty sands moulded to cylindrical specimens at medium relative density. The saturated samples are isotropically consolidated at 100 kPa pressure before shearing. Further, numerical simulations were performed on silty sands by inputting the material parameters into the hypoplastic model. This model requires eight material constants as input including critical friction angle, hardness coefficients, limited void ratios, peak state and stiffness coefficients. These constants were determined for each silty sand combination after conducting basic laboratory tests according to the formulations build in the hypoplastic model program. The experimental trends were compared with numerical model simulations under triaxial testing. The effect of the initial state of soil and the amount of silt fines on the undrained response of fine sands is discussed in detail. The liquefaction susceptibility of silty sand is described based on steady state line concept. The results indicate that the silt sands behave as highly contractive, i.e. more liquefiable when compared with sands.

M. Akhila (✉)

Department of Civil Engineering, SCMS School of Engineering and Technology,
Ernakulam, Kerala, India
e-mail: akhila144@gmail.com

K. Rangaswamy · N. Sankar

Department of Civil Engineering, NIT Calicut, Calicut, India

© Springer Nature Singapore Pte Ltd. 2020

A. Prashant et al. (eds.), *Advances in Computer Methods and Geomechanics*, Lecture Notes in Civil Engineering 56,
https://doi.org/10.1007/978-981-15-0890-5_17



9th World Engineering Education Forum 2019, WEEF 2019

Service Learning in Engineering Education: A Study of Student-Participatory Survey for Urban Canal Rejuvenation in Kochi, India

Sunny George¹, Ratish Menon¹, Pramod Thevanoor¹ and John Tharakan^{2,*}

¹SCMS Water Institute, SSET Campus, Karukutty, Kerala, India

^{2,*}College of Engineering and Architecture, Howard University, Washington DC 20059, USA

Abstract

It is widely accepted that learning through doing, or service learning (SL) and engaging students in community centred project based learning (PBL) is transformative in terms of enhancing student learning and employability, effectively improving both technical and soft skills that are sought after by employers, while at the same time growing and developing an informed and educated citizenry. Participatory learning here is a pedagogical approach in which students involve themselves in a community-based project, which has proven to be more effective than direct lecture based transfer and absorption of knowledge. In this paper, we present a case study from Kochi city in Kerala, India, where undergraduate (UG) engineering students from the environmental engineering (EE) program at SCMS School of Engineering and Technology (SSET) participated voluntarily in the comprehensive survey of a 10.87 km canal running through busy, dense and heavily populated urban area of Kochi City. This Thevara-Perandoor (T-P) canal was a heavily used commercial artery for the city. Unfortunately, the T-P canal is now totally degraded, primarily due to unregulated solid waste dumping and untreated sewage inflows at numerous locations along its course throughout the urban space. The UG students of the CE program at SSET voluntarily came forward to do the study on behalf of Kochi Municipal Corporation (KMC). This partnership, between an academic program and a community based entity, such as a municipal corporation or any other community based entity, establishes a model for integrating meaningful service learning into engineering education. The partnership provided an immense opportunity for the students to implement whatever they had learned in the classroom and doing so by working for the benefit of the community in which they themselves were resident. This paper describes the practices that are being followed in this service learning exercise. The paper also focuses on the impediments as well as the opportunities that exist for both widening and deepening the knowledge domain of the students, while working on the mentioned urban canal survey. The value and impact of the model described through the examined case study is especially important, given that the notion of service learning as a pedagogical approach is gaining momentum in the Indian engineering education sector, and when programs such as *Unnat Bharath Abhiyan* which focuses on and mandates utilizing student voluntary work for rural development are being implemented.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the 9th World Engineering Education Forum 2019.

Keywords: Service Learning; Urban Canal Rejuvenation; Engineering Education; Kochi; Thevara-Perandoor Canal;

Depth Information Enhancement Using Block Matching and Image Pyramiding Stereo Vision Enabled RGB-D Sensor

Sunil Jacob¹, Member, IEEE, Varun G. Menon¹, Senior Member, IEEE, and Saira Joseph, Member, IEEE

Abstract—Depth sensing devices enabled with an RGB camera, can be used to augment conventional images with depth information on a per-pixel basis. Currently available RGB-D sensors include the Asus Xtion Pro, Microsoft Kinect and Intel RealSense™. However, these sensors have certain limitations. Objects that are shiny, transparent or have an absorbing matte surface, create problems due to reflection. Also, there can be an interference in the IR pattern due to the use of multiple RGB-D cameras and the depth information is correctly interpreted only for short distances between the camera and the object. The proposed system, block matching stereo vision (BMSV) uses an RGB-D camera with rectified/non-rectified block matching and image pyramiding along with dynamic programming for human tracking and capture of accurate depth information from shiny/transparent objects. Here, the IR emitter generates a known IR pattern and the depth information is recovered by comparing the multiple views of the focused object. The depth map of the BMSV RGB-D camera and the resultant disparity map are fused. This fills any void regions that may have emerged due to interference or because of the reflective transparent surfaces and an enhanced dense stereo image is obtained. The proposed method is applied to a 3D realistic head model, a functional magnetic resonance image (fMRI) and the results are presented. Results showed an improvement in speed and accuracy of RGB-D sensors which in turn provided accurate depth information density irrespective of the object surface.



Index Terms—Block matching, depth sensing, disparity estimation, image pyramiding, RGB-D sensor, stereo vision.

I. INTRODUCTION

DEPTH sensing is a challenging task and is the core behind numerous indoor and outdoor applications. Various sensing devices have been employed recently for capturing accurate depth information. Precision of captured information is vital for many biomedical applications and helps in the detection and in-depth analysis of diseases like tumors [1]. RGB-D cameras are one of the few preferred modern

sensing systems that capture RGB images along with per-pixel depth information [2]–[4]. Currently available RGB-D sensors capture depth information at reasonable accuracy with good data transfer rates [5]– [6]. The depth range, field of view (FoV) of depth, depth resolution and frame rate of existing RGB-D sensors are presented in table IV. The depth range accuracy depends on the distance of an object in the frame from the sensor. The accuracy of the depth information is the difference between the measured depth data and the actual depth of the reference object. The precision depends on the continuous measurement of the subsequent depth information in static condition. The accuracy and precision of the depth frame of an object also depends on the distance, colour and temperature. RGB-D cameras employ a mechanism in which they project the IR radiation in an irregular pattern of dots using an IR projector and then utilize the IR cameras to detect the infrared light bounced off the object of interest. The required depth information is captured using the offset observed on comparing the projected pattern and the recorded image. Although depth information is captured with relatively good accuracy, these devices suffer from few major limitations.

Manuscript received December 7, 2019; revised January 22, 2020; accepted January 22, 2020. Date of publication January 24, 2020; date of current version April 16, 2020. This work was supported by IEEE EPICS under Grant 2016-12. The associate editor coordinating the review of this article and approving it for publication was Dr. Amitava Chatterjee. (Corresponding author: Varun G. Menon.)

Sunil Jacob is with the Center for Robotics, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: suniljacob@scmsgroup.org).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: varunmenon@ieee.org).

Saira Joseph is with the Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: saira_joseph@scmsgroup.org).

Digital Object Identifier 10.1109/JSEN.2020.2969324

A Novel Spectrum Sharing Scheme using Dynamic Long Short-Term Memory with CP-OFDMA in 5G Networks

Sunil Jacob, Member, IEEE, Varun G Menon, Senior Member, IEEE, Saira Joseph, Member, IEEE, Vinoj P G, Alireza Jolfaei, Senior Member, IEEE, Jibin Lukose and Gunasekaran Raja, Senior Member, IEEE

Abstract—With the rapid increase in communication technologies, shortage of spectrum will be a major issue faced in the coming years. Cognitive radio is a promising solution to this problem and works on the principle of sharing between cellular subscribers and ad-hoc Device to Device (D2D) users. Existing 5G spectrum sharing techniques work as per a fixed rule and are pre-established. Also, recent game theoretic approaches for spectrum sharing uses unrealistic assumptions with very less practical implications. Here, a novel spectrum sharing technique is proposed using 5G enabled bidirectional cognitive deep learning nodes (BCDLN) along with dynamic spectrum sharing long short-term memory (DSLSTM). Joint spectrum allocation and management is carried out with wireless cyclic prefix orthogonal frequency division multiple access (CP-OFDMA). The BCDLN self-learning nodes with decision making capability route information to several destinations at a constant spectrum sharing target, and cooperate via DSLSTM. BCDLN based on time balanced and unbalanced channel knowledge is also examined. With the proposed framework, expressions are derived for the spectrum allocated to multiple sources to obtain their spectrum targets as a variant of the participation node spectrum sharing ratio (PNSSR). The impression of noise when all nodes broadcast with equal spectrum allocation is also investigated.

Index Terms—5G, Artificial Intelligence, Cognitive Nodes, CP-OFDMA, Deep learning, PNSSR, Spectrum sharing

I. INTRODUCTION

THE enormous demand for localized services has forecasted twofold increase in mobile data traffic compared to the existing fixed IP traffic [1], while the cellular networks fails to deliver this growing demand due to restricted bandwidth and shortage of spectrum [2-5]. Device-to-device (D2D) communication has been proposed and integrated in the next generation mobile network as a possible solution to meet this

rising mobile traffic demand. D2D communication allows user to transmit and share cellular data among close proximity users without the involvement of the base stations [6]. Two approaches are employed in D2D to enhance the quality of service: the direct approach and collaborative approach. The cooperative D2D scheme utilizes the cellular user to relay information between base station and end-user to speed up the communication. Decode-and-forward and amplify-and-forward are the prominent relay schemes employed to achieve cooperation among nodes [7]. One of the important factors in multi-hop communication is relay selection. Although various relay selection methods are used to achieve fast data transmission, many of them suffer from limitations in transmission delay, security.

D2D communication happens via the route discovery protocol with increased power savings as base stations are not involved. Once ad-hoc D2D users determine their route, they can share the spectrum with cellular users, i.e. they can operate in the same frequency spectrum as licensed cellular radio network. The existing 5G spectrum sharing has a fixed rule and it is pre-established [8-9]. To overcome the issues, 5G enabled bidirectional cognitive deep learning nodes (BCDLN) along with dynamic spectrum sharing long short-term memory (DSLSTM) is proposed. BCDLN are self-learning proactive and predictive node with decision making capability that creates dynamically adaptable clusters. The joint spectrum allocation and management for 5G access wireless CP-OFDMA communication system with numerous source and multiple destination BCDLN based on time variant and invariant channel knowledge is examined. Each of the BCDLN in the network, sends data to different receivers at a fixed spectrum sharing target and cooperates via DSLSTM with different frequency slots.

Sunil Jacob is with Center for Robotics, SCMS School of Engineering and Technology, Karukatty 683576 India. Email: suniljacob@scmsgroup.org

Varun G Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Karukatty 683576 India. Email: varunmenon@ieee.org (Corresponding Author)

Saira Joseph is with the Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Karukatty 683576 India. Email: saira_joseph@scmsgroup.org

Vinoj P G is with the Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Karukatty 683576 India. Email: vinojpp@scmsgroup.org

Alireza Jolfaei is with the Department of Computing, Macquarie University, Australia. Email: alireza.jolfaei@mq.edu.au

Jibin Lukose is with Center for Robotics, SCMS School of Engineering and Technology, Karukatty 683576 India. Email: jbin2nd@gmail.com

Gunasekaran Raja is with the Department of Computer Technology, Anna University, India. Email: dr_gunasekaran@ieee.org

Received May 10, 2020, accepted May 22, 2020, date of publication May 26, 2020, date of current version June 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2997272

An Adaptive and Flexible Brain Energized Full Body Exoskeleton With IoT Edge for Assisting the Paralyzed Patients

SUNIL JACOB^{1,2}, (Member, IEEE), MUKIL ALAGIRISAMY³,
VARUN G. MENON⁴, (Senior Member, IEEE), B. MANOJ KUMAR⁵, N. Z. JHANJHI⁶,
VASAKI PONNUSAMY⁷, P. G. SHYNU⁸, AND VENKI BALASUBRAMANIAN^{9,10}, (Member, IEEE)

¹Department of Electronics and Communication Engineering, Lincoln University College, Petaling Jaya 47301, Malaysia²Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India³Department of Electrical and Electronics Engineering, Lincoln University College, Petaling Jaya 47301, Malaysia⁴Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India⁵Department of Automobile Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India⁶School of Computer Science and Engineering, Taylor's University, Subang Jaya 47500, Malaysia⁷Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Kampar 31900, Malaysia⁸School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India⁹School of Science, Engineering and Information Technology, Federation University, Mount Helen, VIC 3350, Australia¹⁰Andra Tech Ventures Pvt Ltd., Wyndham Vale, VIC 3024, Australia

Corresponding authors: Sunil Jacob (sunil@scmsgroup.org) and Vasaki Ponnusamy (vasaki@utar.edu.my)

This work was supported in part by the Institute of Electrical and Electronics Engineers (IEEE) EPICS, USA, under Grant 2016-12.

ABSTRACT The paralyzed population is increasing worldwide due to stroke, spinal cord injury, post-polio, and other related diseases. Different assistive technologies are used to improve the physical and mental health of the affected patients. Exoskeletons have emerged as one of the most promising technology to provide

In rehabilitation, the exoskeletons are used to work in parallel with the human legs and carry out the desired actions with ease. These devices are specifically designed to treat disabilities of patients in a clinical setting. The rehabilitation exoskeleton helps paralyzed patients to engage with real-world things and to monitor the movement of body parts. Exoskeletons are also designed for healthy subjects, enabling them to interact with a virtual environment [9]. As healthy people use these haptic exoskeletons, ease of wearability is not a major issue, but portability and efficient finger tracking are highly required. In recent times, to assist children having cerebral palsy disorders, exoskeletons have been designed [10]. The architecture of the control unit, the mechanical system, and feature extraction is discussed in detail. In [11], a wearable hip assist robot is discussed, which is used to improve the gait function and reduce muscle effort and metabolic activities. The device can reduce knee and ankle muscle activity along with a decrease in hip movements. The robot can stabilize the trunk during walking in adults. But the system has not investigated the effectiveness of gait rehabilitation.

In [12], the translation of gait without using crutches gait in a biped robot is demonstrated. The mathematical hybrid model analysis is carried out to find different gait and walking speeds. The walking gaits are stabilized using a centralized

Software-Defined Network (SDN) assisted solutions with exoskeletons for use in rehabilitation are also proposed recently [22]. The majority of the existing exoskeletons have weight, flexibility, and adaptability constraints. Easy wearability and portability are other significant limitations experienced by current assistive exoskeleton-based solutions for rehabilitation [23]–[25].

To overcome the current issues existing with exoskeletons, we propose an adaptive and flexible Brain Energized Full Body Exoskeleton (BFBE) for assisting the paralyzed people. In the BFBE system, the brain signals captured by the EEG sensors are used for controlling the movements of the exoskeleton. The flexibility is incorporated into the system by a modular design approach. The BFBE system has a BCI module, a Control Unit (CU), and a Body-Part Actuation Module (BAM). BCI module captures the brain signal and transforms it into a signal that can be used by the CU. The processing happens at the edge, thus reducing delay in decision making, and the system is further integrated with an IoT module that helps to send an alert message to multiple caregivers in case of an emergency. The system is non-invasive, and the fabricated EEG sensor is used to collect the signals from the scalp. An instrumentation amplifier is used to enhance the strength of the obtained signals. The output signal from the amplifier is subjected to filtering and pre-processing. The signals are generated for different basic human actions (sitting,

1-1-2020

A secure and energy-efficient opportunistic routing protocol with void avoidance for underwater acoustic sensor networks

Varun Menon

DIVYA Midhunchakkaravarthy

SONALI JOHN

SUNIL JACOB

AMRIT Mukherjee

Follow this and additional works at: <https://journals.tubitak.gov.tr/elektrik>



Part of the [Computer Engineering Commons](#), [Computer Sciences Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Menon, Varun; Midhunchakkaravarthy, DIVYA; JOHN, SONALI; JACOB, SUNIL; and Mukherjee, AMRIT (2020) "A secure and energy-efficient opportunistic routing protocol with void avoidance for underwater acoustic sensor networks," *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 28: No. 4, Article 34. <https://doi.org/10.3906/elk-2001-51>

Available at: <https://journals.tubitak.gov.tr/elektrik/vol28/iss4/34>

This Article is brought to you for free and open access by TÜBİTAK Academic Journals. It has been accepted for inclusion in Turkish Journal of Electrical Engineering and Computer Sciences by an authorized editor of TÜBİTAK Academic Journals. For more information, please contact academic.publications@tubitak.gov.tr.

A secure and energy-efficient opportunistic routing protocol with void avoidance for underwater acoustic sensor networks

Varun MENON^{1,2*} , Divya MIDHUNCHAKKARAVARTHY³ , Sonali JOHN² , Sunil JACOB^{4,5} ,
Amrit MUKHERJEE^{6,7} 

¹Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya, Malaysia

²Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Cochin, India

³Center of Postgraduate Studies, Lincoln University College, Petaling Jaya, Malaysia

⁴Department of Electronics and Communication Engineering, Lincoln University College, Petaling Jaya, Malaysia

⁵Center for Robotics, SCMS School of Engineering and Technology, Cochin, India

⁶Department of Computer Science and Communication Engineering, Jiangsu University, Jiangsu, P.R. China

⁷School of Electronics Engineering, KIIT University, Bhubaneswar, India

Received: 11.01.2020

Accepted/Published Online: 04.05.2020

Final Version: 29.07.2020

Abstract: Recently, underwater acoustic sensor networks (UASNs) have gained wide attention due to their numerous applications in underwater surveillance, oil leakage detection, assisted navigation, and disaster prevention. With unique characteristics like increased propagation delay, constant mobility of sensor nodes, high error rate, and limitations in energy and interference, efficient routing of data packets from the source node to the destination is a major challenge in UASNs. Most of the protocols proposed for traditional sensor networks do not work well in UASNs. Although many protocols have been specifically proposed for underwater environments, the aim of most of them is to improve only the quality of service (QoS) in the network. The security of the transmitted data, energy efficiency of the participating nodes, and handling of communication voids are three significant challenges that need to be adequately addressed in UASNs. In this research work, a secure and energy-efficient opportunistic routing protocol with void avoidance (SEEORVA) is proposed. This protocol uses the latest opportunistic routing strategy for reliable data delivery in the network and also provides priority to the nodes having energy above a specific threshold in the forwarding process, thereby increasing the lifetime and energy efficiency in the network. The transmitted messages are encrypted using a secure lightweight encryption technique. The protocol is also integrated with a strategy to handle the communication voids in the network. Simulation results with Aqua-Sim validate the better performance of the proposed system compared to the existing ones.

Key words: Energy efficiency, communication voids, routing protocols, secure data transmission, QoS, underwater acoustic sensor networks

1. Introduction

The ocean covers about 70% of the Earth's surface and is the most abundant source of rare and valuable resources. Due to various constraints, knowledge about the underwater environment is limited, and most of these resources are still unexplored. Underwater acoustic sensor networks (UASNs) [1] have given us hope as a possible solution to this problem. A UASN is a group of self-driven sensor nodes and autonomous vehicles connected underwater to perform different collective tasks based on user applications [2]. Sensor nodes placed at various locations and depths sense and record data and transfer them through the network of nodes to the

*Correspondence: varungmenon46@gmail.com

destination sinks placed at the surface. The collection centers are usually on buoys or ships on the water's surface. Integrating with the most popular Internet of Things (IoT) [3, 4] technology, this smart network of interconnected underwater devices forms the Internet of Underwater Things (IoUT) [5].

Recently, the IoUT and UASNs have gained wide popularity due to their numerous research, industrial, and military applications. They are currently deployed for underwater monitoring and surveillance, oil leakage detection, assisted navigation, and disaster prevention. UASNs are different from the traditional sensor networks (TSNs) and use acoustic signals instead of radio signals. Routing of data packets in UASN is an exceedingly challenging task due to the unique features of the transmission medium such as long propagation delay, constant mobility of sensor nodes, high error rate, limitations in energy, increased error rate, interference caused by animals, and limited bandwidth. Reliable communication and efficient transfer of data from the source to the destination node are vital factors determining the success of various user applications deploying UASNs with multiple objectives. Routing protocols proposed for TSNs do not work well in the underwater environment [6–8]. In the last few years, many techniques have been discussed for efficient data transfer in UASNs, with opportunistic routing protocols (ORPs) [9, 10] being the latest and most efficient among them. ORPs use a broadcasting strategy to increase the number of forwarder nodes and create a prioritized list of available forwarder nodes. They then select the node that has maximum progress to the destination for forwarding the data packet. If that node is unable to forward the data packet within a specified time limit, the next forwarder node in the list forwards the data packet, thus ensuring reliable data delivery in the network. Although the ORPs proposed for UASNs offer several advantages, most of them are designed primarily for improving the quality of service (QoS) in the network.

One of the main limitations in UASNs is the difficulty in periodic recharging of the sensor nodes. If the energy available in the sensor nodes gets exhausted very quickly, the nodes cannot participate in future data transmission. Hence, it is essential to optimize the energy usage in data packet forwarding and conserve energy to extend the lifetime of each sensor node [11–13]. This issue is very inadequately addressed by most of the routing protocols proposed for UASNs [14–16]. Security in data transmission is another major issue to be addressed in UASNs. Sensor nodes in many military and industrial applications collect and record sensitive data. These sensitive data have to be securely stored and transmitted to the sink nodes and any leakage can be very harmful [17]. Moreover, it is found that numerous communication voids [18] occur in underwater environments. Communication voids occur when a source node is unable to find any suitable forwarder node in its transmission range and located in the direction of the destination. Communication voids are also called communication gaps or the unreachability problem. Failure of intermediate nodes due to energy drainage, wrong deployment, intrusions, attacks, etc. are some of the reasons contributing to the occurrence of voids in the network. As most of the latest routing protocols use a position-based greedy forwarding mechanism, this issue has become a major concern. Lack of proper mechanisms to handle voids can lead to huge data loss and loss of energy with retransmissions. Our proposed protocol, the secure and energy-efficient opportunistic routing protocol with void avoidance (SEEORVA), addresses all three issues and also supports good QoS for data transmission in the underwater environment.

In SEEORVA, the sensed and collected data are encrypted using a lightweight security protocol, the novel tiny symmetric encryption algorithm (NTSA) [19]. These encrypted packets are sent to the destination nodes through the network of sensor nodes. Only the collection and processing centers located at the surface are capable of decrypting these data packets, hence ensuring the security of transmitted data. The proposed protocol uses an opportunistic routing strategy but considers the remaining energy in each sensor node as a

significant factor determining the selection of the next best forwarder node. Nodes that have less energy are given less priority to participate in the forwarding process, thus extending the lifetime of each sensor node. SEEORVA is also integrated with a unique strategy to handle the communications voids in the network. Simulation results using Aqua-Sim [20] validated the better performance of SEEORVA compared to the existing protocols in the underwater environment.

The paper is arranged as follows. Section 2 presents the discussion on a few related works. Section 3 discusses the proposed work. Theoretical analysis of energy-efficient data transfer in the network is also presented. Section 4 presents a discussion on the results achieved through simulations. Here the performance of the proposed work is compared with existing protocols in the underwater environment. Section 5 presents the future research directions in UASNs. Finally, the paper concludes in section 6 with future research directions. Table 1 presents the notations and Table 2 presents the abbreviations used in the article. Table 3 lists the differences between TSNs and UASNs.

Table 1. Frequently used notations.

19 Notation	Definition
\bar{x}	Transmission signal coefficient
\bar{h}	Fading signal coefficient
\bar{n}	Noise coefficient
$y_1, y_2 \dots y_L$	Signal received at the receiving node
\bar{y}	Receiving signal coefficient
\bar{w}	Beamforming coefficient
B_f	Beamformer output
P_s	Signal power
P_n	Noise power
N_e	Effective noise at the output of beamformer
SNR_{B_f}	SNR at the output of the beamformer
\bar{w}	Beamforming vector

Table 2. List of abbreviations.

Abbreviation	Description
UASN	Underwater acoustic sensor networks
QoS	Quality of Service
SEEORVA	Secure and energy-efficient opportunistic routing protocol with void avoidance
IoT	Internet of things
IoUT	Internet of underwater things
TSN	Traditional sensor networks
ORP	Opportunistic routing protocols
NTSA	Novel tiny symmetric encryption algorithm
PDR	Packet delivery ratio
PFL	Priority forwarder list

Table 3. Difference between TSNs and UASNs.

TSN	UASN
Sensor nodes are deployed densely	Sparse deployment of sensor nodes
The Communication medium is radio waves	The Acoustic channel is the medium used
Data transfer rate is comparatively high	Data transfer rate is low
Less delay in data transmission	High delay is communication and data transmission
Lower energy consumption	High energy consumption
Higher number of static nodes	Higher number of dynamic nodes
Low error rate	High error rate

2. Related work

This section presents and discusses a few of the existing protocols proposed for UASNs. The security of the transmitted data, energy efficiency of the nodes, and handling of communication voids are three major challenges that need to be adequately addressed in UASNs. Several protocols are proposed to improve the QoS and energy efficiency in UASN. In Su et al. [21], a technique is discussed to increase the network lifetime of the sensor nodes in UASNs using the concept of Deep Q-Network. The technique is also aimed at reducing the delay in data transmission in the network. Another method [22] uses the fuzzy-based relay selection approach to select the node with the maximum energy for forwarding the data packets. Furthermore, the holding time is set for each group of forwarding nodes to avoid collision and save energy. Many protocols work on reducing the collision between the nodes in the network, like the multichannel MAC protocol discussed in Bouabdallah et al. [23]. Although many protocols have tried to improve the energy efficiency and QoS in UASNs, most of them have given very little importance to security in data transmission. A reliable security framework for UASNs is proposed in Ateniese et al. [24]. Common security measures and threats faced in UASNs are discussed in detail in that work. It aims to provide data confidentiality, integrity, and authentication for applications deploying UASNs. A comprehensive discussion on the security attacks faced in UASNs is presented in Shahapur and Khanai [25]. A technique to improve the secrecy of block transmissions based on the half-duplex nature of the underwater transceivers in underwater acoustic channels is presented in Huang et al. [26]. A few protocols are designed to handle the communication voids in the network [27, 28]. However, most of the existing protocols focus primarily on improving the QoS. Lack of an efficient technique for energy efficiency and void avoidance with adequate security in data transmission is still a major problem. The proposed method, secure and energy-efficient opportunistic routing protocol with void avoidance (SEEORVA), uses the latest opportunistic routing strategy for reliable data delivery in the network. The protocol considers only the nodes having energy above a specific threshold in the forwarding process, thereby increasing the lifetime and energy efficiency in the network. The transmitted messages are encrypted using a lightweight encryption technique and the protocol is integrated with a strategy to handle the communication voids in the network. The next section discusses the proposed method.

3. Proposed system

3.1. Theoretical analysis

In this section, the theoretical analysis of the proposed work in efficient energy utilization of a UASN is presented and discussed. Emphasis is given on optimizing the beamforming between the sender and the receiver nodes

such that minimum energy is utilized during the transmission process. Here, assuming \bar{x} as the transmission signal coefficient, \bar{h} as the fading signal coefficient, \bar{n} as the noise coefficient, and y_1, y_2, \dots, y_L as the signal received at the receiving node, the receiving signal coefficient \bar{y} is given by

$$\bar{y} = \bar{h}\bar{x} + \bar{n} \tag{1}$$

Here the beamforming coefficient is assumed to be \bar{w} such that

$$\bar{w} = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ w_l \end{bmatrix} \text{ and } \bar{w}^H = [w_1^* \quad w_2^* \quad w_3^* \quad \dots \quad w_L^*].$$

Combining the received signals with the beamforming coefficient \bar{w} , the beamformer output B_f is obtained as

$$B_f = [w_1^* \quad w_2^* \quad w_3^* \quad \dots \quad w_L^*] \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_L \end{bmatrix} \tag{2}$$

$$B_f = \bar{w}^H \bar{y} \tag{3}$$

Substituting the value of the received output signal \bar{y} in Eq. 1 into Eq. 2, we obtain

$$B_f = \bar{w}^H (\bar{h}\bar{x} + \bar{n}) \tag{4}$$

$$B_f = \bar{w}^H \bar{h}\bar{x} + \bar{w}^H \bar{n} \tag{5}$$

Here the signal power P_s is given by $\bar{w}^H \bar{h}\bar{x}$ and noise power P_n is given by $\bar{w}^H \bar{n}$. Introducing constant P with signal power, we obtain

$$P_s = |\bar{w}^H \bar{h}|^2 . P \tag{6}$$

Now we have the effective noise at the output of beamformer, $N_e = \bar{w}^H \bar{n}$. Calculating the expectation E of N_e we have

$$N_e = E \left\{ |\bar{w}^H \bar{n}|^2 \right\} \tag{7}$$

$$N_e = E \left\{ (\bar{w}^H \bar{n}) (\bar{w}^H \bar{n})^* \right\} \tag{8}$$

where $\bar{w}^H \bar{n} = w_1^* n_1 + w_2^* n_2 + \dots + w_L^* n_L$ and $(\bar{w}^H \bar{n})^* = w_1 n_1^* + w_2 n_2^* + \dots + w_L n_L^*$

$$N_e = E \left\{ \sum_{i=1}^L |w_i|^2 |n_i|^2 + \sum_i \sum_j w_i w_j^* n_i^* n_j \right\}, \tag{9}$$

where $i \neq j$. Here $E(n_i n_j^*) = E(n_i)E(n_j^*)$, which will become zero. Thus the equation is reduced to

$$N_e = \sum |w_i|^2 E\{n_i\}^2 \tag{10}$$

$$N_e = \sigma_n^2 \sum w_i^2 \tag{11}$$

$$N_e = \sigma_n^2 \bar{w}^2 \tag{12}$$

$$N_e = \sigma_n^2 \bar{w}^H \bar{w} \tag{13}$$

Now the *SNR* at the output of the beamformer is given by

$$(SNR)_{\max} = \frac{|\bar{w}^H h|^2 P}{\sigma_n^2 (\bar{w}^H \bar{w})} \tag{14}$$

The aim is to select the beamforming vector \bar{w} , such that \bar{w} maximizes the *SNR* in multiple diversity receiving nodes,

$$(SNR)_{\max} = \left(\frac{|\bar{w}^H \bar{h}|^2}{\bar{w}^H \bar{w}} \right) * \frac{p}{\sigma_n^2} \tag{15}$$

Let us assume that the optimal \bar{w} is K such that

$$(SNR)_{\max} = \left(\frac{K^2 |\bar{w}^H \bar{h}|^2}{K^2 \bar{w}^H \bar{w}} \right) * \frac{p}{\sigma_n^2} \tag{16}$$

Here the constant K will get canceled, implying that scaling will have no effect. It is scale invariant and we need to select \bar{w} such that its magnitude is one. Select \bar{w} such that $\|\bar{w}\|^2 = 1$, which implies $\bar{w}^H \bar{w} = 1$. Now

$$(SNR)_{\max} = |\bar{w}^H \bar{h}|^2 \cdot \frac{p}{\sigma_n^2} \tag{17}$$

Now we need to find the maximum value for $\left(|\bar{w}^H \bar{h}|^2 \frac{p}{\sigma_n^2} \right)$, for which $\bar{w} = \text{const } \bar{h}$. For this we have $c^2 \|\bar{h}\|^2 = 1$, which gives $C = \frac{1}{\bar{h}}$. Optimal beamforming vector \bar{w} that maximizes the received *SNR* = $\frac{\bar{h}}{\bar{h}}$ = \bar{w} (the optimal value that is calculated as maximum ratio combiner). Now we have $SNR = \left(\frac{(\bar{h}^H \bar{h})^2}{\bar{h}} \frac{p}{\sigma_n^2} \right)$, which gives $SNR = h^2 \frac{p}{\sigma_n^2}$. This is the optimal *SNR* at the output of the receiver.

3.2. Best forwarder selection

The proposed method uses the opportunistic routing strategy for the selection of the best relay node. The opportunistic routing strategy enables the presence of more than one forwarder node and hence the chances of data delivery at the destination are high. Thus, our algorithm initially makes sure that the packet delivery ratio (PDR) in the network is high and maximum packets are delivered at the destination. The proposed algorithm used for the best forwarder selection is presented below. At first, the source node that has to transmit data packets creates a virtual vector pipe to the destination node. A list of nodes that are located within the pipe is then compiled. The highest energy of the nodes in the list is then calculated. The nodes that are outside the pipe are not considered in the forwarding process. A threshold energy value based on the calculated highest energy value is set for the forwarder nodes. The node that has energy above the threshold and within the transmission range of the source node and also that has the maximum progress to the destination is chosen as the best forwarder node. If this node cannot forward the packet within a specific period (set by a timer) due to reasons like mobility or damage of the node, the next node in the list forwards the data packet to the destination. Thus, a high rate of data delivery along with energy efficiency is guaranteed by the proposed approach. The working of this technique is illustrated in Figure 1. Here the source ‘s’ wants to transmit the data packets to target node ‘t’. The source node creates a list ‘f’, ‘c’, ‘h’, ‘a’, ‘b’ of the nodes that are located within the vector pipe. Now the source node calculates the highest energy node among the four and sets the threshold value. Nodes ‘h’, ‘c’, and ‘f’ are within the transmission range of the source node. Therefore, a priority forwarder list (PFL) (‘h’,‘c’,‘f’) is generated by the source node based on the maximum progress to the destination. Let us assume that node ‘h’ does not have energy greater than the set threshold. Thus, the node ‘c’ is selected as the best forwarder. If the node ‘c’ is unable to forward the data packet within a particular time, node ‘f’ forwards it.

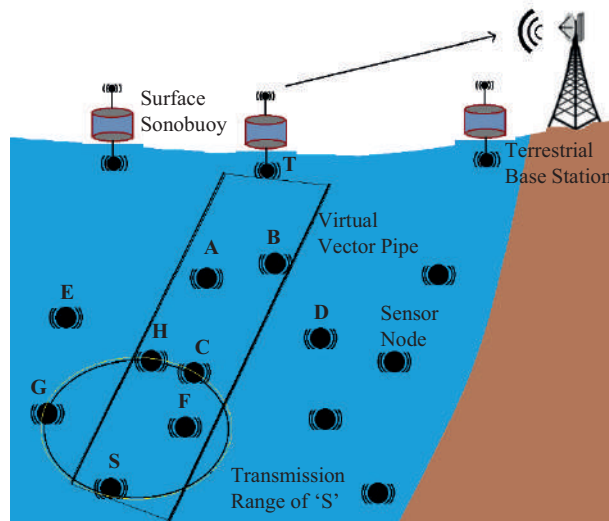


Figure 1. Illustration of the best forwarder selection algorithm using the proposed method.

Algorithm 1: For best forwarder node selection

1. The source node creates a virtual vector pipe to the destination node.
2. The source node checks for the nodes inside the pipe. If true go to step 3, otherwise drop the data packet.
3. Calculate the highest energy among the nodes inside the pipe and go to step 4

4. Set the energy threshold as highest/2.
5. The source node creates a PFL containing the nodes within its transmission range.
6. Sort the list by the distance to the destination. The node having a minimum distance to the destination is assigned the highest priority in the PFL.
7. Check whether the energy $>$ threshold for all the nodes in the PFL. If true keep the node in the list and go to step 9. Otherwise, go to step 8.
8. Drop the packet and label as a low energy node.
9. Call packet forwarding algorithm.
10. Repeat the step until the packet reaches the destination.

Algorithm 2: For packet forwarding

1. Data packet is received by a source node.
2. The node generates the PFL.
3. Forward the packet to the nodes in the PFL.
4. If the best priority forwarder forwards the data packet within a period, go to step 6, else go to step 5.
5. Next node in the PFL forwards the data packet.
6. The received node repeats steps from 1 to 4 until the packet reaches the destination.

3.3. Security and void avoidance in data transmission

One of the major issues to be addressed in UASNs is the security of data transmitted between the sensor nodes. In many UASN applications such as military ones, the security of data is the most important factor. Any leakage of information in such applications can have major consequences. To provide high security for the transmitted data, we integrate a simple, lightweight, and strong encryption technique discussed in Rajesh et al. [19] into our proposed system. It is a symmetric encryption algorithm that follows the Feistel structure. The algorithm has 64 rounds and 32 cycles of operation. In every cycle, there is an odd and even round. The message to be transferred through the network is set as 64-bit blocks. The key used in the algorithm is 128 bits and is divided into 4 subkeys. The subkeys are then dynamically applied to odd and even rounds in encryption. This algorithm is used to encrypt the transmitted data packet, which can only be decrypted by the receiver, thereby protecting the data from intruders. Due to sparse deployment and frequent mobility of sensor nodes, communication voids are a major issue contributing to increased packet drops. An efficient mechanism to handle communications voids is necessary to guarantee good QoS for various applications in UASNs. In the proposed method when a node experiences a communication void, it sends a data packet void_alert to the previous node from which the data have come. The previous node tries to find an alternate route avoiding the void or around the void and routes the data packets to the destination. All the remaining packets through the void node are redirected to this new route until the void node reports that the void is past. This technique gives much better results compared to the major existing techniques used to handle communication voids. The major advantage of this technique is that it is easy to implement with less overhead and delay.

4. Results and discussion

The performance of the proposed method, SEEORVA, is analyzed and compared with that of the existing underwater routing protocols using simulations in Aqua-Sim [20]. Aqua-Sim is an extended version of NS-2 and offers easy implementation of underwater network scenarios. The specifications used for the simulation are presented in Table 4.

The protocols are compared in terms of performance using the packet delivery ratio (PDR) (number of packets delivered at the destination compared to the total packets sent), average end-to-end delay, and normalized energy consumption. The PDR and average end-to-end delay are used to measure the QoS in the network. The number of nodes participating in the network is varied and the performance of the protocols is measured. Comparison is done using vector-based forwarding (VBF) [29] and vector-based void avoidance (VBVA) [30] protocols. Figure 2 presents the comparison of PDR with a varying number of nodes. Here we can see that the proposed technique has a better PDR compared to all the existing techniques with different numbers of nodes. Moreover, in void scenarios the proposed technique achieves a good PDR. This is because, in the proposed technique, the best forwarder is selected using an opportunistic strategy and even if the best forwarder is unable to forward the data packet, the next best forwarder forwards it, thus ensuring reliability and a high packet delivery rate in the network. Figure 3 presents the comparison of average end-to-end delay with varying number of nodes. SEEORVA has less delay in data transmission compared to all other existing techniques like VBF and VBVA. The proposed technique is easy to implement with less complexity. The efficient forwarder priority list in the algorithm aids in less delay in data transmission. Figure 4 presents the normalized energy consumption achieved through various protocols with varying numbers of nodes in the network. It is evident that SEEORVA has better energy efficiency in the network compared to the existing approaches. In addition, the proposed technique is tested in a communication void environment. Both in normal and void scenarios, the proposed technique achieves much better energy efficiency compared to all the existing approaches. This is because the proposed technique considers the residual energy available in the sensor nodes for the forwarder selection process. Thus, the network lifetime is extended with improved energy efficiency in the routing process. The simulation results show the better performance offered by the proposed technique in terms of QoS, energy efficiency, and security compared to the existing techniques in UASNs.

Table 4. Simulation specifications.

Parameter name	Values
Simulator name	NS 2.35 with Aqua-Sim
Dimension of topology	1500 x 1500 x 1500 m
Transmission range	250 m
Antenna-Type	Omni-Directional
Data rate	50 kbps
Packet size	25 to 125 bytes
Number of nodes	100 to 300
Simulation time	200 s
Number of Simulation runs	10
Protocols	SEEORVA, VBF, VBVA

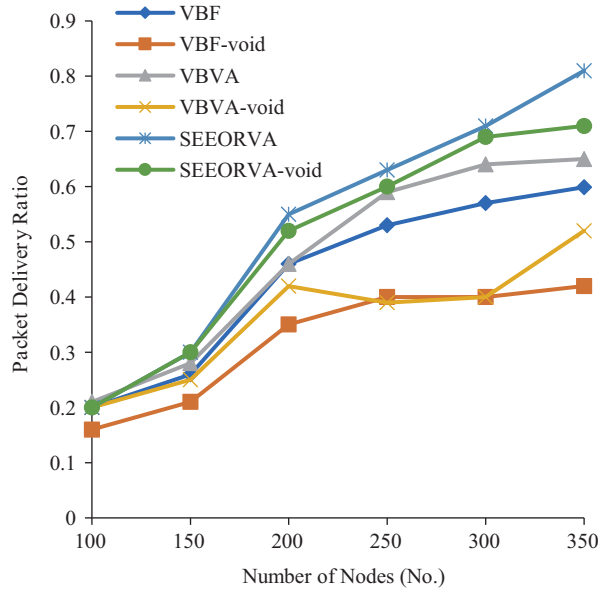


Figure 2. Variation in packet delivery ratio (PDR) with different numbers of nodes.

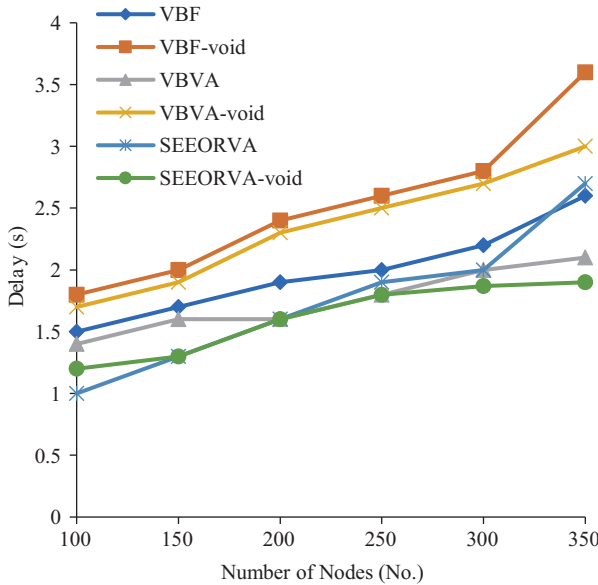


Figure 3. Variation in delay with different numbers of nodes.

5. Future research directions

In this section, we list a few research areas within UASNs that have generated interest among researchers due to the opportunities, issues, and challenges.

- Quality of service: QoS has been one of the major research areas focused on in UASNs. As the success of most of the applications depends on a high delivery rate, less delay, and other QoS parameters, numerous studies have been carried out in this direction. With an increased number of sensor nodes and advancement in sensor technology, new technique for further optimization of QoS in the network is an emerging area of research.
- Energy efficiency: With restrictions and limitations in recharging the deployed sensor node underwater,

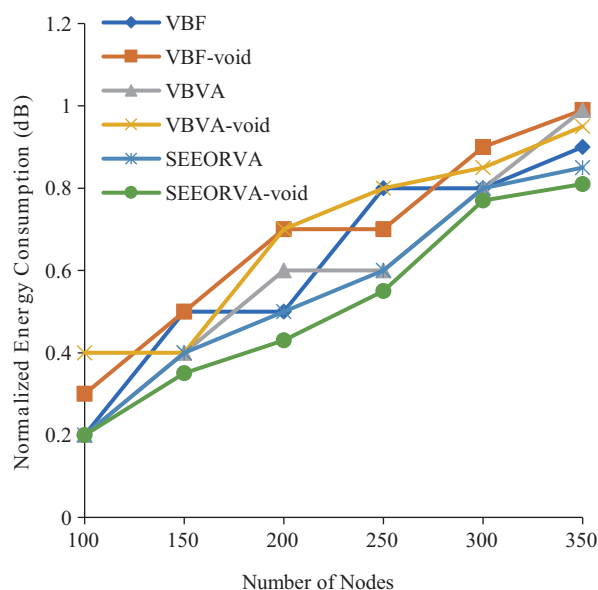


Figure 4. Variation in normalized energy consumption with different numbers of nodes.

the major hindrance behind the success of many applications is related to energy efficiency. Current research focuses on optimizing energy usage in the routing process and preventing energy leakages. This will be a major area of research in the future too.

- **Channel utilization:** Efficient utilization of the channel is another major area of research in UASNs that has gained wide prominence. With numerous challenges like propagation delay, constant mobility of sensor nodes, high error rate, and interference, it is vital to have optimal utilization of the channel.
- **Security:** The security of data transmitted between sensor nodes is a major area of concern. In many UASN applications such as military ones, the security of data is the most important factor. Any leakage of information in such applications can have major consequences. Recently, numerous studies have been carried out to secure the communication between the sensor nodes in UASNs. With increasing attacks and threats, research in security and privacy will be an ongoing and highly challenging task.
- **Reliability:** Many studies have focused on reliable data delivery in the network. This is an important parameter because it provides trust for user applications and helps in its success.
- **Communication voids:** Sparse deployment and frequent mobility of sensor nodes have led to increased communication voids in networks, leading to frequent packet drops. An efficient mechanism to handle communications voids is necessary to guarantee good QoS for various applications deployed with UASNs. This is a major research area in UASNs and will continue to be prominent in the coming years.

6. Conclusion

The security of the transmitted data, energy efficiency of the nodes, and handling of communication voids are three major challenges in UASNs that are not adequately addressed in most of the existing protocols. To address these issues, a secure and energy-efficient opportunistic routing protocol with void avoidance (SEEORVA) is proposed. This protocol uses the latest opportunistic routing strategy for reliable data delivery in the network and considers only the nodes having energy above a specific threshold in the forwarding process, thereby

increasing the lifetime and energy efficiency in the network. The transmitted messages are encrypted using a lightweight encryption technique and the protocol is also integrated with a strategy to handle the communication voids in the network. Simulation results with Aqua-Sim confirmed the better performance of the proposed system compared to the existing ones. The proposed technique needs to be tested in a real-time underwater environment in the future. The design could also be modified to incorporate bulk data coming from numerous sources.

References

- [1] Jouhari M, Ibrahim K, Tembine H, Ben-Othman J. Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things. *IEEE Access* 2019; 7: 96879-96899. doi: 10.1109/ACCESS.2019.2928876
- [2] Menon V. Opportunistic routing protocols in underwater acoustic sensor networks: issues, challenges, and future directions. In: Fei Hu (editor). *Magnetic Communications: From Theory to Practice*. USA: CRC Press, 2017, pp. 127-148.
- [3] Ang K, Seng J. Application specific Internet of Things (ASIoTs): taxonomy, applications, use case and future directions. *IEEE Access* 2019; 7: 56577-56590. doi: 10.1109/ACCESS.2019.2907793
- [4] Anavangot V, Menon V, Nayyar A. Distributed big data analytics in the internet of signals. In: *Proceedings of the IEEE International Conference on System Modelling and Advancement in Research Trends (SMART)*; Moradabad, India; 2018. pp. 73-77. doi: 10.1109/SYSMART.2018.8746983
- [5] Liou EC, Kao CC, Chang CH, Lin YS, Huang CJ. Internet of underwater things: challenges and routing protocols. In: *Proceedings of the IEEE International Conference on Applied System Invention*; Chiba, Japan; 2018. pp. 1171-1174. doi: 10.1109/ICASI.2018.8394494
- [6] Menon V, Joeprathap PM. Comparative analysis of opportunistic routing protocols for underwater acoustic sensor networks. In: *Proceedings of the IEEE International Conference on Emerging Technological Trends (ICETT)*; Kollam, India; 2016. pp. 1-5. doi: 10.1109/ICETT.2016.7873733
- [7] Alzubi JA, Almomani O, Alzubi OA, Al-shugran M. Intelligent and dynamic neighbourhood entry lifetime for position-based routing protocol using fuzzy logic controller. *International Journal of Computer Science and Information Security* 2016; 14(1): 118-128.
- [8] Zorzi M, Casari P, Baldo N, Harris A. Energy-efficient routing schemes for underwater acoustic networks. *IEEE Journal on Selected Areas in Communications* 2008; 26(9): 1754-1766. doi: 10.1109/JSAC.2008.081214
- [9] Coutinho RWL, Boukerche A. Opportunistic routing in underwater sensor networks: potentials, challenges and guidelines. In: *Proceedings of 2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*; Ottawa, Canada; 2017. pp. 1-2. doi: 10.1109/DCOSS.2017.42
- [10] Menon V, Joeprathap PM. Analysing the behaviour and performance of opportunistic routing protocols in highly mobile wireless ad hoc networks. *International Journal of Engineering and Technology* 2016; 8(5): 1916-1924. doi: 10.21817/ijet/2016/v8i5/160805409
- [11] Irandoost A, Taheri S, Movaghar A. PL-MAC: ProLonging network lifetime with a MAC layer approach in wireless sensor networks. In: *Proceedings of the Second International Conference on Sensor Technologies and Applications*; Cap Esterel, France; 2008. pp. 109-114. doi: 109-114.10.1109/SENSORCOMM.2008.120
- [12] Alrabea A, Alzubi OA, Alzubi JA. A task-based model for minimizing energy consumption in WSNs. *Energy Systems* 2019. <https://doi.org/10.1007/s12667-019-00372-w>
- [13] Almomani O, Al-Shugran M, Alzubi JA, Alzubi OA. Performance evaluation of position-based routing protocols using different mobility models in manet. *International Journal of Computer Applications* 2015; 119(3): 43-48.
- [14] Kaur K, Kumar N, Garg S, Rodrigues JPC. EnLoc: data locality-aware energy-efficient scheduling scheme for cloud data centers. In: *Proceedings of the IEEE International Conference on Communications (ICC)*; Kansas City, USA; 2018, pp. 1-6. doi: 10.1109/ICC.2018.8422225

- [15] Kaur K, Garg S, Kaddoum G, Gagnon F, Kumar N et al. An energy-driven network function virtualization for multi-domain software defined networks. In: Proceedings of the IEEE INFOCOM - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); Paris, France; 2019. pp. 121-126. doi: 10.1109/INFOCOMW.2019.8845314
- [16] Kaur K, Garg S, Kaddoum G, Bou-Harb E, Choo KR. A big data-enabled consolidated framework for energy efficient software defined data centers in IoT setups. *IEEE Transactions on Industrial Informatics* 2020; 16(4): 2687-2697. doi: 10.1109/TII.2019.2939573
- [17] Jiang S. On securing underwater acoustic networks: A survey. *IEEE Communications Surveys and Tutorials* 2019; 21(1): 729-752. doi: 10.1109/COMST.2018.2864127
- [18] Menon V, Joeprathap PM. Opportunistic routing with virtual coordinates to handle communication voids in mobile ad hoc networks. In: Proceedings of the Second International Symposium on Signal Processing and Intelligent Recognition Systems; Trivandrum, India; 2015. p. 323-334.
- [19] Rajesh S, Paul V, Menon V G, Khosravi M. A secure and efficient lightweight symmetric encryption scheme for transfer of text files between embedded IoT devices. *Symmetry* 2019; 11(2): 293. doi:10.3390/sym11020293
- [20] Xie P, Zhou Z, Peng Z, Yan H, Hu T et al. Aqua-Sim: An NS-2 based simulator for underwater sensor networks. In: Proceedings of the OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges; Biloxi, USA; 2009. pp. 1-7. doi: 10.23919/OCEANS.2009.5422081
- [21] Su Y, Fan R, Fu X, Jin Z. DQELR: An adaptive deep Q-network-based energy and latency-aware routing protocol design for underwater acoustic sensor networks. *IEEE Access* 2019; 7: 9091-9104. doi: 10.1109/ACCESS.2019.289159
- [22] Mulla A, Jadhav V. Energy efficient routing protocol for underwater acoustic sensor network. In: Proceedings of the Second International Conference on Intelligent Computing and Control Systems; Madurai, India; 2018. pp. 1707-1712. doi: 10.1109/ICCONS.2018.8662952
- [23] Bouabdallah F, Zidi C, Boutaba R, Mehaoua A. Collision avoidance energy efficient multi-channel MAC protocol for underwater acoustic sensor networks. *IEEE Transactions on Mobile Computing* 2019; 18(10): 2298-2314. doi: 10.1109/TMC.2018.2871686
- [24] Ateniese G, Caposelle A, Gjanci P, Petrioli C, Spacciniet D. SecFUN: Security framework for underwater acoustic sensor networks. In: Proceedings of OCEANS 2015; Genova, Italy; 2015. pp. 1-9. doi: 10.1109/OCEANS-Genova.2015.7271735
- [25] Shahapur S, Khanai R. Localization, routing and its security in UWSN - A survey. In: Proceedings of the International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT); Chennai, India; 2016. pp. 1001-1006. doi: 10.1109/ICEEOT.2016.7754836
- [26] Huang Y, Xiao P, Zhou S, Shi Z. A half-duplex self-protection jamming approach for improving secrecy of block transmissions in underwater acoustic channels. *IEEE Sensors Journal* 2016; 16(11): 4100-4109. doi: 10.1109/JSEN.2015.2446465
- [27] Sonali J, Menon V, Nayyar A. Simulation-based performance analysis of location-based opportunistic routing protocols in underwater sensor networks having communication voids. In: Proceedings of Data Management, Analytics and Innovation; New Delhi, India; 2019. pp. 697-711.
- [28] Wang Z, Han G, Qin H, Zhang S, Sui Y. An energy-aware and void-avoidable routing protocol for underwater sensor networks. *IEEE Access* 2018; 6: 7792-7801. doi: 10.1109/ACCESS.2018.2805804
- [29] Peng X, Cui J, Lao L. VBF: vector-based forwarding protocol for underwater sensor networks. In: Proceedings of the International conference on research in networking; Coimbra, Portugal; 2006. pp. 1216-1221. doi: 10.1007/11753810_111
- [30] Peng X, Zhou Z, Peng Z, Cui JZ, Zhi Z. Void avoidance in three-dimensional mobile underwater sensor networks. In: Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications; Boston, MA, USA; 2009. pp. 305-314. doi: 10.1007/978-3-642-03417-6_30



SysDroid: a dynamic ML-based android malware analyzer using system call traces

Ananya A.¹ · Aswathy A.¹ · Amal T. R.¹ · Swathy P. G.¹ · Vinod P.¹ · Mohammad Shojafar^{2,3}

Received: 10 August 2019 / Revised: 29 November 2019 / Accepted: 31 December 2019 / Published online: 13 January 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Android is a popular open-source operating system highly susceptible to malware attacks. Researchers have developed machine learning models, learned from attributes extracted using static/dynamic approaches to identify malicious applications. However, such models suffer from low detection accuracy, due to the presence of noisy attributes, extracted from conventional feature selection algorithms. Hence, in this paper, a new feature selection mechanism known as *selection of relevant attributes for improving locally extracted features using classical feature selectors* (SAILS), is proposed. SAILS, targets on discovering prominent system calls from applications, and is built on the top of conventional feature selection methods, such as mutual information, distinguishing feature selector and Galavotti–Sebastiani–Simi. These classical attribute selection methods are used as local feature selectors. Besides, a novel global feature selection method known as, weighted feature selection is proposed. Comprehensive analysis of the proposed feature selectors, is conducted with the traditional methods. SAILS results in improved values for evaluation metrics, compared to the conventional feature selection algorithms for distinct machine learning models, developed using Logistic Regression, CART, Random Forest, XGBoost and Deep Neural Networks. Our evaluations observe accuracies ranging between 95 and 99% for dropout rate and learning rate in the range 0.1–0.8 and 0.001–0.2, respectively. Finally, the security evaluation of malware classifiers on adversarial examples are thoroughly investigated. A decline in accuracy with adversarial examples is observed. Also, SAILS recall rate of classifier subjected to such examples estimate in the range of 24.79–92.2%. However, prior to the attack, the true positive rate obtained by the classifier is reported between 95.2 and 99.79%. The results suggest that the hackers can bypass detection, by discovering the classifier blind spots, on augmenting a small number of legitimate attributes.

Keywords Android malware · Machine learning (ML) · Deep learning (DL) · Feature selection · Adversarial machine learning (AML) · Attacks

✉ Mohammad Shojafar
mohammad.shojafar@unipd.it; m.shojafar@surrey.ac.uk

Ananya A.
ananya@scmsgroup.org

Aswathy A.
aswathy@scmsgroup.org

Amal T. R.
amal@scmsgroup.org

Swathy P. G.
swathy@scmsgroup.org

Vinod P.
vinodp@scmsgroup.org

¹ Department of Computer Science & Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala, India

² ICS/5GIC, University of Surrey, Guildford GU27XH, UK

³ University of Padua, 35131 Padua, Italy

1 Introduction

The number of Android users has exponentially increased over the past decade, and this has opened the doors for the attackers to innovate methods, to compromise devices through vulnerable Android applications. The rise in the number of malware variants, constraint the anti-virus vendors in the signature update process, thereby adversely affecting the security of smartphones and tablets. According to Symantec's Internet Security Threat Report 2018, there was an increase of 54% Android malware variants with the figures reported in 2017 [9]. Symantec's report also mentions that third-party app stores hosted 99.9% of malicious apps. Kaspersky Lab products detected 5,321,142 malicious installation packages, 151,359 new mobile banking Trojans and 60,176 new mobile ransomware Trojans for the year 2018 [22].

Conventional mechanisms of Android malware analysis are based on three approaches which are static, dynamic and hybrid analysis. The static analysis incorporates signature-based, permission-based, and component-based investigation. Dynamic analysis involves the execution of the application in real-time and observing the behaviour of the application. The hybrid analysis consists of combining static and dynamic features [24]. In literature, several researchers have developed machine learning (ML) techniques [17–19] to resolve the problems of Android malware attacks. Previous works have focused on evasion attacks, which results in misclassification of the sample [6]. Zhou and Jiang [37] have worked on detecting Android malware by AV vendors. DroidAPIMiner [1] performs extraction of API call frequency from android applications and performs malware detection using supervised learning algorithms.

Contributions Static analysis alone cannot be used to identify malware applications as malware hide payload in the encrypted form, installed on execution. To cope with this problem, it is essential to develop methods based on dynamic analysis. Hence, some questions arises: Is it possible to adopt strategies employing dynamics analysis to detect Android malware using different features? How can the proposed dynamic method be guaranteed, to swiftly update itself to dynamically changing real-time Android samples (applications)? The main aim of the paper is to respond to these queries by devising a new feature selection method, and evaluate the robustness of the classifier against adversarial attacks. Hence, a novel system call analysis is proposed to detect Android malware at run time. In this way, a new feature selection method called *SAILS* is proposed, which improves the performance of classifiers over the conventional feature selection methods. The classical identified by us in this paper are mutual

information (MI), Galavotti–Sebastiani–Simi (GSS) and Distinguishing Feature Selector (DFS), to extract relevant attributes representative of the target class. Experiments are conducted on benchmark dataset consisting of 2474 Drebin malware and 2475 benign apps. In summary, the main contributions of our work are listed as follows:

- A new feature selection mechanism known as Selection of relevant Atttributes for Improving Locally extracted features using classical feature Selectors (SAILS) is proposed.
- An extensive analysis of ML and deep learning (DL) algorithms under diverse classifier parameters is conducted.
- One of the key observation is that XGBoost has a higher prediction capability in comparison to other classification algorithms.
- The performance of the classification algorithms when subjected to adversarial examples is performed. It is experimentally verified that classifiers are misled even to small modification in attributes introduced by augmenting malware samples with few prominent benign features.

The rest of the paper is organised as follows. In Sect. 2, the related works proposed in the field of Android malware detection. Section 3 presents the methodology. The attack model is presented in Sect. 4. The experiments, results, and its analysis has been discussed in Sect. 5. Finally Sect. 6 discusses about the conclusion and future work.

2 Related work

Several anti-malware techniques have been introduced to detect malware on Android devices. These techniques can be broadly classified as static and dynamic analysis. In static analysis, malicious behaviour is analyzed by scanning the source code of the application, instead of executing an application/program. The source code of the program is investigated to identify the trigger of malicious event. On disassembling the apps, different features such as permissions, hardware components, intents, broadcast receivers, data flow, APIs, control flow, etc. can be derived. Dynamic analysis is performed during run time. Here, the malware scanners monitor the response generated by the operating system, on the execution of the program. Commonly used features include network connections, system calls, etc. To bridge the gap between static and dynamic analysis, hybrid anti-malware techniques were also developed to improve the performance of malware detector. In the following subsections, the background is categorized into static methods presented in Sect. 2.1, solutions based

on dynamic malware analysis approach presented in Sect. 2.2, and hybrid techniques presented in Sect. 2.3.

2.1 Static feature based approaches

The authors in [11] propose a method to evaluate the security of machine learning-based android malware detector. Classifiers were trained using API calls extracted from the smali files. To explore the security of the classifier, the authors considered evasion attacks on diverse threat models. They propose a robust secure-learning paradigm that can be applied to other security tasks like fraud detection. The goal of the research work was to improve the robustness of online malware scanners against adversarial examples created at test time.

Another group of authors in [14] designed a method that worked in discrete and binary input domains. For malware detection, they train the neural network on the Drebin dataset and achieve classification performance against similar works in literature. They use an AML algorithm to mislead the ML model to about 63% of all malware samples in the Drebin dataset.

2.2 Dynamic feature selection methods

Bhandari et al. [5] proposed a malware detection tool capable of handling code injection during runtime. This approach binds semantics of the program and a classification engine. A sequence of system calls demonstrates the semantics of the app. To capture the actual behaviour of the apps, the order of the system calls were conserved. They apply Markov property on the acquired system call traces and construct a sequential system call graph (SSG). The authors compute all the acyclic paths by considering the first, and the last system calls as the start and end node in SSG. They develop feature vectors from the typical sets by applying asymptotic equipartition property on each path. After that, they perform statistical analysis by finding the average logarithmic branching factor of each path to train the model. Further, they use the histogram binning technique to form the feature vector table and train them using supervised learning algorithms. They conduct experiments on dataset containing 2000 applications (1000 Benign and 1000 malware apps). Benign apps were collected from the Google play store, and malware apps belonged to 119 different families. The proposed system obtains a detection accuracy of 94.2%.

The authors in [12] have designed a two-step learning strategy named *KuafuDet*, that used adversarial detection to learn malware patterns. It is composed of an offline training step that could select and extract features from the samples and further use this model to compromise the online tool. An automated camouflage detector is used to

filter the false negatives and feed them back into the training phase. *KuafuDet* reduces false negatives and improves detection accuracy by 15%. This method was tested on more than 250,000 mobile applications to demonstrate the scalability of *KuafuDet*.

Additionally, in another work, the authors [35] propose a method that transforms the packed malware variant detection problem, to a system calls classification problem. They generated a sequence of sensitive system calls and further applied principal component analysis to extract relevant attributes. Then, multi-layer neural networks were utilized to classify benign and malicious applications. The proposed system was reported to achieve a detection accuracy of 95.6%.

The authors in [8] used strace tool to extract the system calls. Each invocation of system calls was mapped to a frequency-based feature vector. Experiments were performed on self-made 60 apps, and the proposed work reported accuracy between 85 and 100%. Later authors, in [2] proposed an ML-based dynamic malware detection method. They extracted API calls and system call traces. In particular, 74 API functions and 90 system calls were considered as features. For classification, Random Forest classifier was utilized. Experiments were conducted on the dataset comprising of 7520 apps, of 3780 samples were used for training and remaining (3740 application) were considered in the testing phase. The study demonstrated an accuracy of 96.66%.

The authors in [10] introduced an ML approach that helped in increasing the user effectiveness in handling system data to improve security and privacy. The proposed approach was evaluated on different ML algorithms deployed on real-world systems, and it showed better efficiency. Suciu et al. [29] developed a Fail Attacker Model. The model effectiveness was evaluated on adversaries having limited capabilities. A poisoning attack was subjected on different ML algorithms. Consequently, the fail model exhibited better resilience on generalized transferability.

2.3 Hybrid feature selection methods

The authors in [25] propose a novel malware detection method called *HADM*—hybrid analysis for detection of malware. The set of features were represented as vectors fed to Deep Neural Network (DNN) with different kernels. To apply graph kernels onto the graph sets, they converted dynamic information into graph-based representations. Further, output from various vector were combined with graph feature sets using hierarchical multiple kernel learning (MKL), to build a final hybrid classifier. The authors in [4] propose a three-level hybrid model called *SAMADroid* for Android malware detection. By

combining the benefits of three levels, SAMADroid provides more detection accuracy on static and dynamic analysis, deployed on local and remote host. For static analysis, attributes from the manifest files and the dex code files were considered. For dynamic analysis, they generate system call logs based on the local inputs and send it to the remote server. Remote servers analyze the extracted static features and the logs. Finally, ML algorithms are used to classify malware and benign apps. They present that the SAMADroid can provide detection accuracy of 82.76% with Random Forest (RF). Also, the authors in [30] designed a novel hybrid approach and utilized the NetLink technology to generate patterns of system calls related to file and network access. They compare the pattern of malware and benign apps, to build a malicious pattern set. They aim to collect system calls and in offline compares both the malicious and normal pattern sets to identify the unknown app.

Besides, the paper [15] explains a hybrid method called *MalDAE* that combines the dynamic and static API sequences into one hybrid sequence based on semantics mapping. *MalDAE* experiments show that the system achieved accuracy in the range of 94.39% and 97.89%. Also, *MalDAE* provides an overall idea about the common types of malware and predictive support for understanding and resisting malware. Another novel hybrid method is [32] which consists of deep Autoencoder (DAE) and convolutional neural network (CNN). They reconstruct high-dimensional features of Android applications and use multiple CNN to detect Android malware. To increase sparseness, nonlinear activation function *Relu* was utilized to prevent overfitting in the serial CNN architecture *CNN-S*. To increase the capability of feature extraction, they bound the convolutional layer and pooling layer with the full-connection layer. Later, deep Autoencoder as a pre-training method of CNN was employed to reduce the training time. They tested their hybrid approach on 10,000 legitimate and 13,000 malicious apps. This resulted in 5% increase in the accuracy compared with SVM. Further, training time using the DAE-CNN model reduces by 83% compared with the CNN-S model. A hybrid scheme designed by Zhenlong Yuan et al. [34] aims to introduce an online deep-learning-based malware detection method called *DroidDetector*. *DroidDetector* performs prediction on thousands of Android apps and also thoroughly perform an in-depth study of the features and declares an accuracy of 96.76%.

3 Methodology

In this section, our proposed system is presented. Figure 1 describes the architecture and the working of our system. A clear description of the steps involved in the identification and evaluation of malware samples, is discussed in the subsequent sections.

3.1 Data collection

The data collection phase involves the collection of two types of Android applications, malware applications as well as trusted applications. A total of 4949 samples comprising of 2475 benign and 2474 Drebin applications [3, 23] were downloaded. In particular, legitimate application were considered from diverse app categories. Categories comprised of lifestyle, education, medical, comics, etc. from 9Apps site [38]. All these apps were scanned using VirusTotal [31], which is a web service that examines files or URLs to check whether they are malicious or not. Finally, samples labelled as benign were included for carrying out experiments.

3.2 Feature extraction

A system call sequence presents how an application requests a service from the kernel of the operating system. System calls generated during the execution of an application, are used as features to identify samples as malware or benign. The intention of using system calls is to deduce the behaviour of the application and to understand its interaction with the Android operating system. System calls have been logged using strace utility. Further, to mimic human interactions, Android Monkey Runner [20], a utility in sdk is periodically accessed. Specifically, Android Monkey was configured to direct 200 random UI events in a minute. Some events generated are: (i) reception of SMS (ii) answer and make call, (iii) change geolocation, (iv) swipes and (v) update the battery charge status These set of events were selected as it generalizes the operations performed on the smartphones. To collect the call traces, the following procedures were adopted:

1. Install the app (using ADB install command)
2. Extract the package name and class of the application
3. Gets the process id (PID) corresponding to each application
4. Invoke strace command, and at each timestamp log system calls
5. Start Monkey command with application package name as the parameter
6. Suspend the application for ten seconds
7. Kill the app

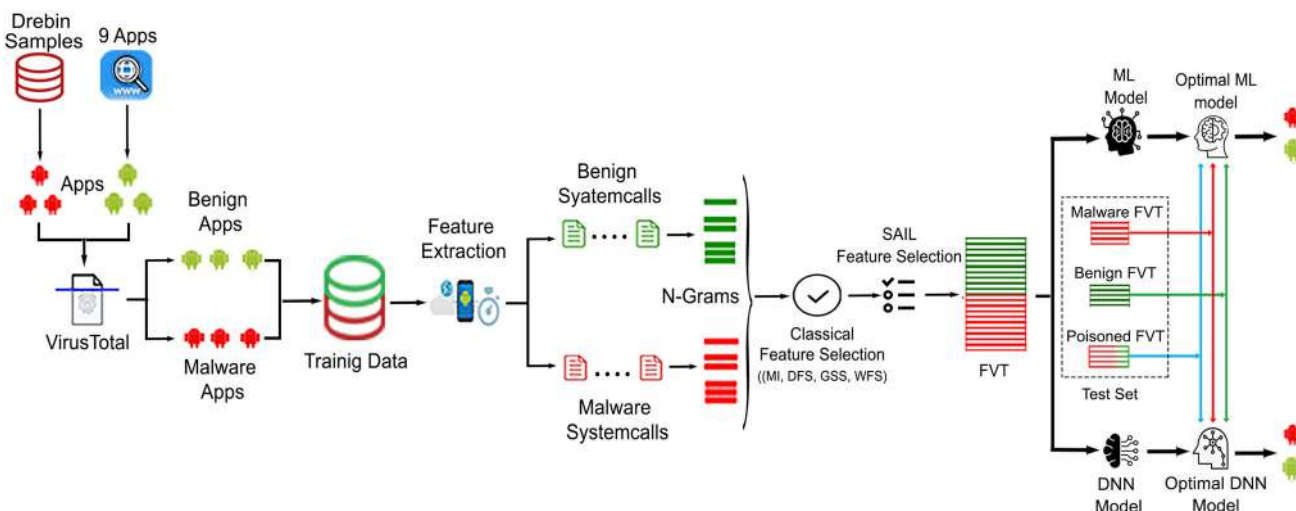


Fig. 1 The proposed architecture

8. Uninstall the app (using ADB uninstall command)
9. Delete app from device, finally reset the emulator clean state

Alternatively, in-order to extract system calls a human expert well versed in Android programming also interacted with the installed applications. In the feature extraction phase, initially, an emulator or an Android virtual device is created by specifying the basic configuration details such as device name, memory size, OS version, storage area, skins, screen resolution etc. Subsequently the .apk file, is installed into the emulator using the command: `adb -s emulator-id install sample_app.apk`.

Likewise, all other applications on the emulator are installed, and the system calls (`strace -p process_id -o output_path`) are recorded. Simultaneously, the Android monkey continuously interacts with the app during (`adb shell monkey -p pkg_name -v 200`). Once the specified events are completed, the process stops, and finally, the emulator is restored back into its clean state.

3.3 N-gram generation

In this part of the paper, the process of generation of *N-grams* is discussed. *N-gram* is a sequence of n items from a given sample. In *N-gram* model, the occurrence of an item is predicted based on the occurrence of its previous $n - 1$ items. They are used to store the context of the words and can be used to make the next word predictions. *N-gram* for any range usually perform the best, and is shown to be applied in the domain of malware detection [27]. Since *N-grams* overlap, they not only capture the statistics about substrings of length but also implicitly capture frequencies of longer substrings as well.

Experiments on unigrams, bigrams, and trigrams are performed. Consider for an example a set of extracted features and the corresponding unigrams, bigrams, and trigrams generated:

- (i) **Features:** read prctl openat epoll_ct socketpair recvform
- (ii) **Unigram:** read, prctl, openat, epoll_ct, socketpair, recvform
- (iii) **Bigram:** read:prctl, prctl:openat, openat:epoll_ct, epoll_ct:socketpair, socketpair:recvform
- (iv) **Trigram:** read:prctl:openat, prctl:openat:epoll_ct, openat:epoll_ct:socketpair, epoll_ct:socketpair:recvform

Once the features are obtained, the corresponding bigrams and trigrams are generated, and analysis is performed on all *N-grams* to investigate the effect of *N-gram* size on classification accuracy.

3.4 Feature selection

Feature selection is one of the most crucial phase of ML, which has a huge impact on the classification model generated. Irrelevant and redundant features must be removed or else they may negatively impact the classification. Irrelevant are variables/attributes which have less correlation with a class, and redundant features have correlation with one or more attributes in the feature space. Thus, feature selection allows us to filter out attributes so that only the important features which help in classification is left out. Some of the conventional feature selection methods previously applied in the domain of malware detection are MI [21], GSS [36] and DFS [33]. Mutual information (MI) is a measure between two random variables quantifying the amount of information obtained from one random variable compared to the other random variable. DFS method exploits the hypothesis that certain term

appearing in more number of documents are highly relevant for categorization. GSS coefficient is a simplified variant of the Chi-square statistics proposed by Galavotti et al. In this approach, Chi-square is used to measure the correlation between the attribute and class. Thus, if a variable carries more information about a target class, such attributes are known as a characteristic feature, also have a high value. Thus, in summary, the benefits of performing feature selection improve accuracy reduce overfitting, and reduces training time. The following subsection highlights the feature selection methods employed in our experiment.

3.4.1 SAILS: selection of relevant attributes for improving locally extracted features using classical feature selectors

We propose *SAILS* as a novel feature selection method. The following steps outline the proposed feature selection methods through a simple example.

Step 1 Set of malware and benign system calls are listed. Let S_i denote a system call.

	S_1	S_2	S_3	S_4
M	bind	munmap	capget	ioctl
B	fcntl	lseek	writew	prctl

Step 2 Feature selection algorithms such as MI, GSS and DFS are used to give score for the samples.

Let $FS = \{f_1, f_2, \dots, f_n\}$
 $f_i(a_j) = \text{Malware/Benign score}$

Step 3 System calls are sorted based on the malware and benign scores.

	S_1	S_2	S_3	S_4
M	bind	munmap	capget	ioctl
B	fcntl	lseek	writew	prctl

Sorted list containing system calls is arranged in descending order of prominence.

M	S_2	S_4	S_3	S_1
B	S_1	S_4	S_2	S_3

Step 4 System calls are added to the new list

To do so, first, system calls from both malware and benign list are taken. Then, a check is performed to identify whether the system call already exists in the final list, if it is not present, it is augmented to the final result.

M	S_2	S_4	S_3	S_1
B	S_1	S_4	S_2	S_3

S_2	S_1		
-------	-------	--	--

Next, the second system call from both malware and benign list is picked and their presence in the final list is

checked. Here S_4 appears in both malware and benign list, hence it is added once in the final list.

M	S_2	S_4	S_3	S_1
B	S_1	S_4	S_2	S_3

S_2	S_1	S_4	
-------	-------	-------	--

Similarly, the third system call is chosen, i.e., S_3 and S_2 . Here, S_2 is already present in the final list, hence only S_3 is added in the final list.

M	S_2	S_4	S_3	S_1
B	S_1	S_4	S_2	S_3

S_2	S_1	S_4	S_3
-------	-------	-------	-------

Here, both S_1 and S_3 are already present in the final list, hence no need to add these two system calls again.

M	S_2	S_4	S_3	S_1
B	S_1	S_4	S_2	S_3

S_2	S_1	S_4	S_3
-------	-------	-------	-------

This method is built over the conventional feature selection methods such as MI, GSS, and DFS. Initially, the malware and benign score of the union of malware and benign system calls are computed using the conventional feature selection methods. Then the features are ranked in the descending order of scores for the target classes. Once two separate lists of system calls are obtained, the new list of attributes based on SAILS are derived. For this purpose, the alternate system calls are selected, one each from the malware score ranked list and the other from the benign score ranked list and these system calls, are added into the new list provided, they are not the same and are not already present in the list. If the two calls are similar, only one instance of the feature is added to the list.

Algorithm 1 gives a brief description of the steps involved in SAILS. The input to the algorithm is the list of system calls, $S = \{s_1, s_2, \dots, s_N\}$; where m is the set of malware system calls and b is the set of benign system calls (line 1). In lines 2–3, call present in the applications are sorted. Lists, u and v consist of system calls arranged in the descending order of precedence of malware and benign score, respectively. X is the set of system calls that forms the final output. In the following steps, it is first checked whether the current sorted malware system call from u is present in X , if not, then, the presence of the benign system call from v is checked in X . If absent, both malware and benign system calls are added to X at positions j and $j + 1$ respectively (lines 7-8). Similarly, if the benign system call is already found in X then only the malware system call is added (line 10). Alternatively, if the malware system call is already present in X , then the presence of benign system call is checked, if not present, only the benign call is added

into X . Finally, if the benign system call from v is already present in X then there is no need to add either the malware (u) or the benign (v) system call into X . Finally, the list X is returned as output, as shown in line number 20.

The implementation of SAILS was performed using binary Max-Heap tree. Thus, two heap trees one for malware and another for benign system calls is obtained. Since this tree is populated using identical (union) system calls from both the dataset, hence it contains the same number of system calls, but arranged in different fashion due to the difference in local scores, corresponding to each system call. In this way, the space complexity in worst case is $O(N)$, to be precise as the number of system calls are less (approximately 393 in Linux Kernel 3.7), the space complexity is also less. The time complexity to create Max-Heap tree is $O(\log N)$. Generally, both the trees contain system call with maximum score at the root node. A system call with maximum score is picked and appended to the list X . During this course of action the system call with the highest score is deleted from the tree and a heapify operation is performed. Thus, the time required to undertake heapify operation in the worst case requires $O(N \log N)$. Therefore, in the worst case the total time to arrange the system calls will need $O(N \log N)$.

Algorithm 1 System calls extracted using SAILS

```

1: Procedure SAILS( $S, m, b$ )
2:  $u \leftarrow \text{sort}(m)$ 
3:  $v \leftarrow \text{sort}(b)$ 
4: while  $i \neq |S|$  do
5:   if  $u[i] \notin X$  then
6:     if  $v[i] \notin X$  then
7:        $X[j] \leftarrow u[i]$ 
8:        $X[j + 1] \leftarrow v[i]$ 
9:     else
10:       $X[j] \leftarrow u[i]$ 
11:    end if
12:  else
13:    if  $v[i] \notin X$  then
14:       $X[j] \leftarrow v[i]$ 
15:    else
16:      //do nothing
17:    end if
18:  end if
19: end while
20: return  $X$ 
    
```

3.4.2 Weighted feature selection (WFS)

Another method used in our experiment for feature selection is WFS. Here the weight of system calls are first computed, and then ordered in descending order of the corresponding malware and benign scores separately. The prominent system calls are used to create the feature vector matrix, which is further used to train the model and to evaluate the performance of the generated model.

Algorithm 2 discusses the steps involved in feature selection using WFS. The input to the algorithm is the set of system calls, S , as shown in line number 1. From line number 3 to 6, we compute the weight of system calls in the malware set. In particular, the weight of system call is computed as the product of the ratio of occurrences of system calls in malwares to the total occurrence of system calls in both training samples and the frequency of the calls in malware files, to the total number of malware samples in the training set. The time complexity for computing the weight is $O(1)$. Similarly, line number 8 to 10 depicts the calculation of the weight of system calls for benign examples. Finally, the average weight of system calls is determined (refer to line 12). The steps 3–12 steps are repeated until the weight corresponding to each call in set S is ascertained.

Algorithm 2 System calls extracted using WFS

```

1: Procedure WFS( $S$ )
2:  $S = \{s_1, s_2, \dots, s_N\}$ ; set of system calls.
3: while  $i \neq |S|$  do
4:   //Determine weight of system call in malware set
5:    $T(S_i, M) \leftarrow \frac{\text{occ}(S_i, M)}{\text{occ}(S_i, M) + \text{occ}(S_i, B)}$ 
6:    $U(S_i, M) \leftarrow \frac{\text{freq}(S_i, M)}{|M|}$ 
7:    $wt(S_i, M) \leftarrow T(S_i, M) * U(S_i, M)$ 
8:   //Determine weight of system call in Benign set
9:    $T(S_i, B) \leftarrow \frac{\text{occ}(S_i, B)}{\text{occ}(S_i, M) + \text{occ}(S_i, B)}$ 
10:   $U(S_i, B) \leftarrow \frac{\text{freq}(S_i, B)}{|B|}$ 
11:   $wt(S_i, B) \leftarrow T(S_i, B) * U(S_i, B)$ 
12:   $\text{avg}(S_i) \leftarrow \frac{wt(S_i, M) + wt(S_i, B)}{2}$ 
13: end while
14: return  $\text{avg}(S_i)$ 
    
```

3.5 Feature vector table

The feature vector table is the collection of vectors consisting of n rows and $m + 1$ columns. Here, n represents the number of applications in the dataset, and m is the number of unique system calls invoked by both malware and

	V_1		V_k		V_m	Y
X_1	a_{11}	...	a_{1k}	...	a_{1m}	0
X_2	a_{21}	...	a_{2k}	...	a_{2m}	0

X_j	a_{j1}	...	a_{jk}	...	a_{jm}	1

X_n	a_{n1}	...	a_{n2}	...	a_{nm}	1

Fig. 2 Feature vector table

benign APKs. An example of a feature vector table is depicted in Fig. 2. The columns 1 to m denote the attributes obtained after the feature selection phase. The last column denotes the class label associated with each sample. Here, the class label ‘1’ identifies malware samples, and ‘0’ denotes benign apps. Each row denotes a vector X with dimension m that correspond to samples in the dataset. Lastly, a_{ij} denotes the number of times j th system call was invoked by i th sample.

3.6 Training and testing phases

Once the feature selection phase is concluded, the next step is to train the model and predict new samples. Training phase largely include construction of a feature occurrence matrix, a data structure that record the frequency of attributes obtained after the feature selection process. Eventually, the classification algorithm learns patterns discriminant to the target classes. Later, the test samples are supplied to a trained ML model. The learned model assigns class labels to each apk in the test set. The performance of the classifiers is evaluated using performance metrics, such as accuracy, precision, recall, and F1-score. Models are developed using Logistic Regression (LR), Classification and Regression Tree, RF, XGBoost, and DNNs.

Train-test split method is used in this work. Specifically, 60% samples are assigned to train set and the remaining 40% of the apks are included in the test set. The train-test split method may sometimes result in the overfitted model.

3.7 Classifiers

Classification or predictive modelling is a method of approximating a hypothesis function (f) which maps to discrete output variables (y) for input observations (X). In particular, classification is a supervised approach where a program learns patterns from input examples and predict the class for a new sample. Multiple ML classification algorithms such as RF [7], Classification and Regression Tree (CART) [26], LR [16] and XGBoost [13] were used in our work. DNNs were also used to analyze the performance and to compare the performance obtained with machine learning algorithms.

3.7.1 Classification and regression tree (CART)

A decision tree is a non-parametric ML technique for regression and classification problems. Given the input observation, decision tree forms a hierarchical structure. Each internal node corresponds to attribute and leaf node corresponds to class labels. CART is a Gini index-based

method. Initially, all training samples are put in the root node. Subsequently, the best partition is explored to minimize the Gini impurity. Noisy or impure attributes classify a randomly selected sample into the wrong subset. Besides, gini impurity equals zero if samples belong to one class. The best discovered partition is further divided into parts, each of which is subsequently seen as a new node. This process is repeated until leaf nodes are obtained.

3.7.2 Random forest (RF)

Random Forest (RF) consists of a large number of decision trees which can function as an ensemble. Each tree is created from the set of a randomly selected subset of training examples. The individual tree generates a class prediction, further, the class that receives maximum votes is the outcome of the entire classification process.

RF hyper-parameters are used to improve the model’s predictive ability. Commonly used hyper-parameters include a number of trees that the algorithm builds before taking the maximum vote or considers the average prediction. A higher number of trees generally increases performance and makes predictions more stable, however, suffers from speed. Alternatively, Random Forest can be configured with another hyperparameter like split criterion, min/max number of leaf nodes, the height of the tree etc.

3.7.3 Logistic regression (LR)

Logistic Regression is used when the target variable (i.e., dependent variable) is a categorical/binary response. A sigmoid function is used as a logistic function which outputs real value for the corresponding input feature vector. The obtained output value is subsequently converted to binary based on the threshold, in particular, the output is the estimated probability. Additionally, the coefficients also help in predicting the importance of each input variable.

3.7.4 XGBoost

XGBoost is a scalable and precise implementation of gradient boosting, developed solely for improving model performance and speed. Gradient Boosting is an ensemble learner, it creates a final model based on a prediction obtained from the collection of individual models. As the predictive power of individual models is weak and susceptible to overfitting, hence ensembles of weak models improve the overall result. Newly generated models can predict error of prior models. XGBoost utilizes gradient descent approach to reduce the error while combining the models

3.7.5 Deep neural network (DNN)

A DNN is a network contain multiple layers in between the input and output layer. In DNN, nodes in each layer are trained on a set of features of the previous output layer. As the number of layers increases, the complexity of the network increases, and it learns complex attributes. At higher dimensions, traditional ML algorithms such as LR, KNN, etc. exhibit poor performance, whereas in the case of neural networks, as the size of the data increases, the performance of the model increases. Additionally, hyperparameters like drop out, and the learning rate may be altered to improve the model performance.

Drop out is a regularization technique for neural networks, to avoid overfitting. It is an approach in which randomly selected neurons are ignored during training. Thus, in the forward pass, the contribution to the activation of the neurons in subsequent layers are ignored, and any updates in weights are also not applicable in the backward phase. Dropout enables a model to learn more robust features. Learning rate indicates the amount of change made to the model during each step of the search process.

3.8 Evaluation of the classifiers

In this paper, evaluation metrics considered are accuracy (see Eq. (1)), precision (see Eq. (2)), recall (or true positive rate) (see Eq. (3)) and F1-score (see Eq. (4)) to identify the classifier performance. *True negative* (α), is the number of truly classified benign samples. *True positive* (β) is the number of correctly classified malware files. The number of mis-classified legitimate applications are referred as *false positive* (γ). Malicious applications wrongly classified as benign are called as *False negative* (θ). Using TP (β), TN (α), FP (γ) and FN (θ), accuracy, recall, precision and F1-score is computed.

$$\text{Accuracy}(\mathcal{A}) = \frac{\alpha + \beta}{\alpha + \beta + \gamma + \theta} \quad (1)$$

$$\text{Precision}(\mathcal{PRC}) = \frac{\beta}{\beta + \gamma} \quad (2)$$

$$\text{Recall}(\mathcal{REC}) = \frac{\beta}{\beta + \theta} \quad (3)$$

$$\text{F1 - Score}(\mathcal{F}_1) = 2 * \left(\frac{\mathcal{PRC} * \mathcal{REC}}{\mathcal{PRC} + \mathcal{REC}} \right) \quad (4)$$

4 Attack model

Adversarial machine learning involves techniques, where the malicious samples injected with attributes of legitimate applications, forces machine learning system to misclassify such perturbed malware apps. The modified(perturbed) examples are also known as adversarial examples.

Adversarial examples can be broadly classified as (a) poisoning attack-performed during the training phase and (b) evasion attack-perturbed samples created in the prediction phase to mislead detection. Additionally, an adversary may use one/more, threat models. These threat model are related to the knowledge an attacker possesses, with reference to a machine learning system. In a white-box threat model, an adversary has complete knowledge of training samples and classifier parameters. Such attack models are used to evaluate the performance of a machine learning system in the worst case. Moreover, in a black-box attack model, the adversary does not have access to classifier and the training set. She can make a limited number of attempts to fool the classification system.

4.1 Evasion attack

In evasion attack, an attackers feed adversarial input to the classifier to increase misclassification. To accomplish this task and adversary create synthetic malware samples, imitating the properties of the benign applications. To start with, 10% (247 apps) of the total malware samples (2474 samples) are chosen. Then, prominent system calls invoked by benign samples absent in malware applications are appended to it at a varied concentration (1%, 2%, 3%, etc.). Such modified malware samples form the test set, which are later used to predict the performance of the models.

Algorithm 3 Evasion attack

```

1: Procedure Poisoning ( $M, S, j$ )
2:  $V \leftarrow j(M)$  //extract  $j$  percentage of samples from
   malware set
3:  $X \leftarrow S(M)$  //percentage of Unique system calls present
   only in benign set are extracted
4: for each file in  $V$  do
5:   append  $X$  at the end of file
6: end for
7: return file

```

Algorithm 3 gives a brief description of the evasion attack. The input to the algorithm includes M , which is the set of malware files and $S = \{s_1, s_2, \dots, s_N\}$, a set of prominent benign system calls which are absent in malware files, to be injected into the malware samples, at the specified concentration for creating adversarial examples. In particular, system calls in benign Max-Heap tree, are referred to as prominent calls which are absent in the first half of the malware Max-Heap tree. Thus, in the worst-case half the number of nodes in malware Max-Heap tree (half the height of tree) is searched. Hence, in worst case the time complexity to search legitimate system call in malware Max-Heap require $O(\log N)$, where N is the total number of system calls. In line 2, $j\%$ of the malware samples from original malware set are extracted. To be

precise, j is the percentage of malware files that needs to be poisoned (line 1). Next, prominent system calls invoked by benign APKs at the required concentration, i.e. (1%, 2%, ...) are selected (line 3). In lines 4 to 7, for each in file V , the system calls from X are added to samples in V depending on the supplied concentration rate. Finally, the output is a set of perturbed malware instances, injected with a sequence of the calls from benign apps (line 7).

5 Performance analysis

In this section, the experiments along with results are detailed. The experiments are conducted on Ubuntu 18.04 platform with the support of Intel Core i5-8250U CPU @ 1.80 GHz with 8GB RAM. The dataset consisted of 2475 benign and 2474 malware applications. Benign applications are downloaded from 9Apps site and verified for benignness by scanning samples using VirusTotal. The apps from Drebin dataset constituted malware set. This dataset contain malware samples belonging to 179 families. An automated tool, i.e., Android Monkey which is a part of the Android SDK, along with `strace` utility are used to record the system calls. On the completion of system calls extraction, the emulator process is killed and clean snapshot of the emulator is loaded for the analysis of subsequent samples. Comprehensive experiments were performed for evaluating the following:

- Performance of machine learning classifiers using proposed feature selection methods.
- Investigation of optimal feature category and feature length.
- Performance comparison of machine learning and deep learning algorithms.
- Investigation of the effect of drop out and learning rate in the performance of deep learning neural network.

5.1 Experiment-I: feature selection method

A novel feature selection method (SAILS) was developed on the top of classical feature selection approaches. Traditional feature selection methods like MI, GSS and DFS were employed as local feature selectors to estimate the score of each system call. These calls were further ranked using our proposed attribute ranking approach, yielding an enhanced outcome, both in terms of evaluation metrics and feature-length.

Table 1 depicts the comparison of accuracies obtained with proposed feature selectors and the conventional approaches. In this experiment, classification models were developed using RF, CART, LR and XGBoost. On observing the results, it is noted that the accuracy obtained

for our proposed system is higher in most of the cases, and for remaining experiments, the performance was at least equal to the conventional approach. It was observed that the highest accuracy was recorded for RF with unigrams, and for bigrams, the best outcome was achieved using XGBoost. In the case of trigrams, LR exhibited the highest accuracy. It is worth mentioning that the results obtained with the proposed methods exceed with fewer feature-length.

5.2 Experiment II: robustness of N-grams

Figure 3 illustrates the malware and benign scores of 20 system calls which were predominant in benign samples. From Fig. 3a, for unigrams, it can be seen that scores are identical or marginally differ. This suggests statistical similarity in the feature vectors for both malware and benign examples. Considering Fig. 3b and c it seen that the benign system call score is lesser when compared to the malware call score. Similar trend is observed in case of GSS^* , WFS^* and MI^* as shown in Appendix Figs. 11, 12 and 13 respectively.

5.3 Experiments-III: comparison of classification algorithms

In this experiment, the effectiveness of different ML algorithms in identifying malware applications is evaluated. In particular LR, Classification and Regression Tree (CART), Random Forest (RF), eXtreme Gradient Boost (XGBoost) and DNN are chosen. Additionally, a comprehensive analysis on the performance of DNN is performed by determining optimal values of dropout and learning rate.

From Fig. 4 it is clear that for GSS feature selection, the accuracy and F1-Score obtained with SAILS is better compared to the conventional approach.

Further, the average score of each system call was computed, the system calls were then arranged in the descending order of the average scores. Finally, sorted calls using the aforesaid approach was utilized in the training phase. For unigram, it was observed that Random Forest achieved the highest accuracy of 95.85% with F1-score of 95.87%. In the case of bigram, XGBoost showed better results with an accuracy of 99.4% and F1-score of 99.4%. Figure 4c depicts that for trigram, LR shows the highest accuracy of 99.34% and F1-score of 99.34%, which was higher than other classifiers.

A similar pattern of results was obtained with Random Forest and XGBoost employing DFS and MI feature selectors as depicted in Figs. 5 and 6. In the case of DFS feature selector, with Random Forest, the highest accuracy and F1-score obtained for unigram was 96.31% and

Table 1 Performance of classification algorithms with proposed feature selection approach and classical feature selectors

Classifier	Feature selectors	Unigram		Bigram		Trigram	
		FL	\mathcal{A}	FL	\mathcal{A}	FL	\mathcal{A}
LR	MI*	100(98)	97.66	100(2212)	97.16	90(18318)	99.44
	AVG-MI	100(98)	91.05	100(2212)	96.41	90(18318)	96.75
	GSS*	50(47)	88.58	80(1769)	99.13	80(16283)	99.34
	AVG-GSS	90(88)	88.43	100(2212)	99.08	100(20354)	99.34
	DFS*	100(98)	91.96	90(1990)	97.61	100(20354)	99.34
	AVG-GSS	90(88)	88.43	100(2212)	99.08	100(20354)	99.34
	WFS*	40(37)	89.3	100(2212)	98.0	100(20354)	99.4
CART	MI*	100(98)	96.5	100(2212)	96.65	100(20354)	98.83
	AVG-MI	100(98)	92.26	100(2212)	94.94	100(20354)	94.78
	GSS*	90(88)	92.67	80(1769)	97.77	80(16283)	98.58
	AVG-GSS	90(88)	92.57	80(1769)	95.08	100(20354)	98.17
	DFS*	90(88)	92.32	70(1548)	96.7	90(18318)	98.22
	AVG-DFS	90(88)	92.02	90(1990)	95.39	90(18318)	97.89
	WFS*	40(37)	93.1	100(2212)	96.7	100(20354)	98.7
RF	MI*	100(98)	96.75	80(1769)	97.21	100(20354)	98.23
	AVG-MI	100(98)	96.1	80(1769)	97.21	100(20354)	96.8
	GSS*	90(88)	95.85	90(1990)	98.78	80(16283)	98.17
	AVG-GSS	100(98)	95.5	100(2212)	98.53	100(20354)	97.77
	DFS*	100(98)	96.31	100(2212)	97.66	100(20354)	97.82
	AVG-DFS	100(98)	96.11	100(2212)	97.21	100(20354)	97.97
	WFS*	90(88)	97.2	20(442)	97.6	100(20354)	97.5
XGBoost	MI*	100(98)	95.6	90(1990)	99.3	100(20354)	99.2
	AVG-MI	100(98)	95.5	90(1990)	99.3	100(20354)	99.2
	GSS*	80(78)	95.75	70(1548)	99.44	80(16283)	98.32
	AVG-GSS	80(78)	92.65	70(1548)	97.67	80(16283)	97.72
	DFS*	90(88)	95.7	100(2212)	99.4	80(16283)	98.2
	AVG-DFS	90(88)	95.4	100(2212)	97.3	80(16283)	98.2
	WFS*	70(68)	96.6	40(884)	97.9	20(4070)	97.7

FL denotes feature length at which best outcomes were obtained. FL is represented in the form of $P(Q)$, where P denote the percentage of features extracted from the feature space and Q denote the number of attributes used to create model

Asterisks indicate a revised feature set after the application of SAILS

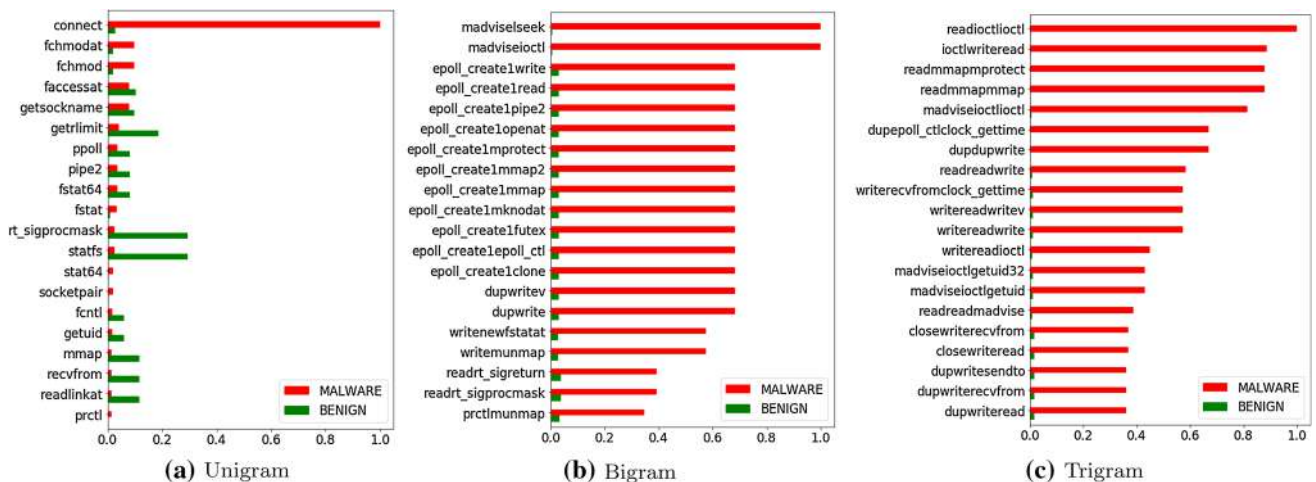


Fig. 3 Score difference of N-grams for DFS*

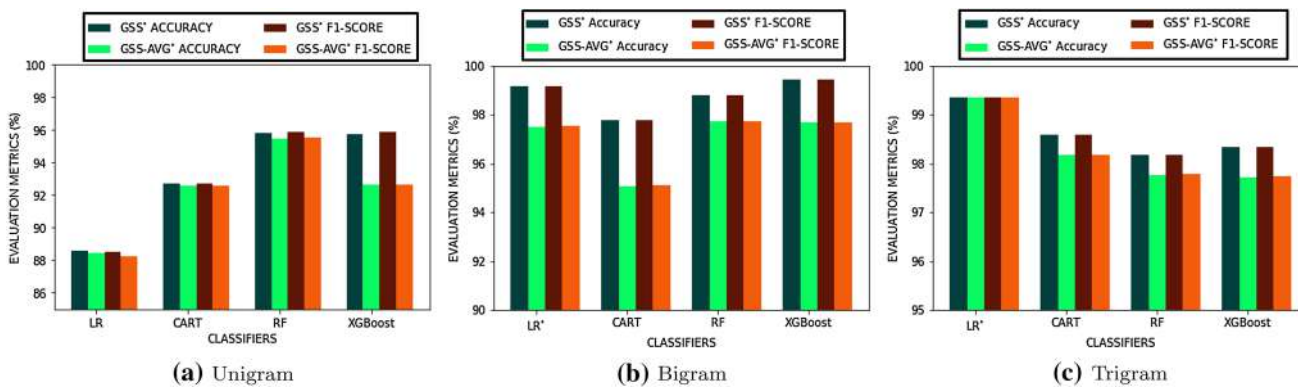


Fig. 4 Performance of GSS* feature selection on unigram, bigram, trigram

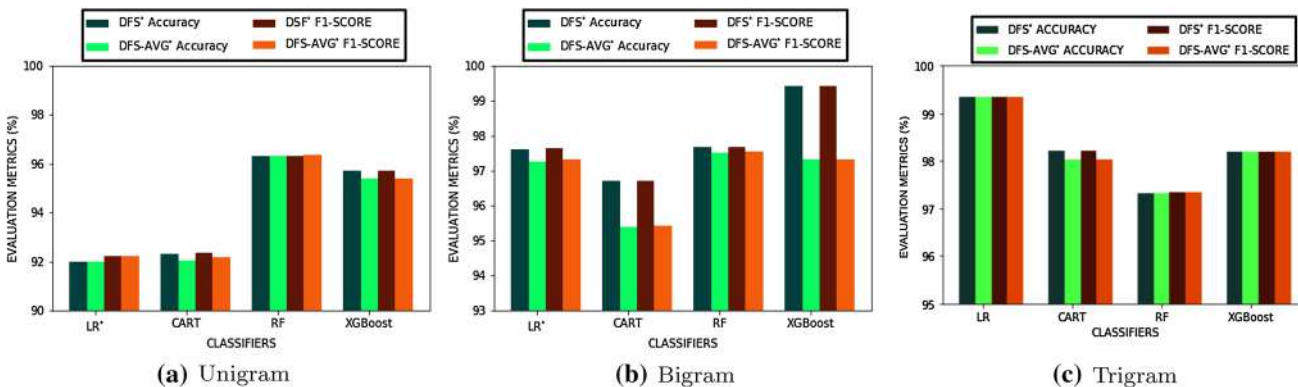


Fig. 5 Performance of DFS* feature selection on unigram, bigram, and trigram

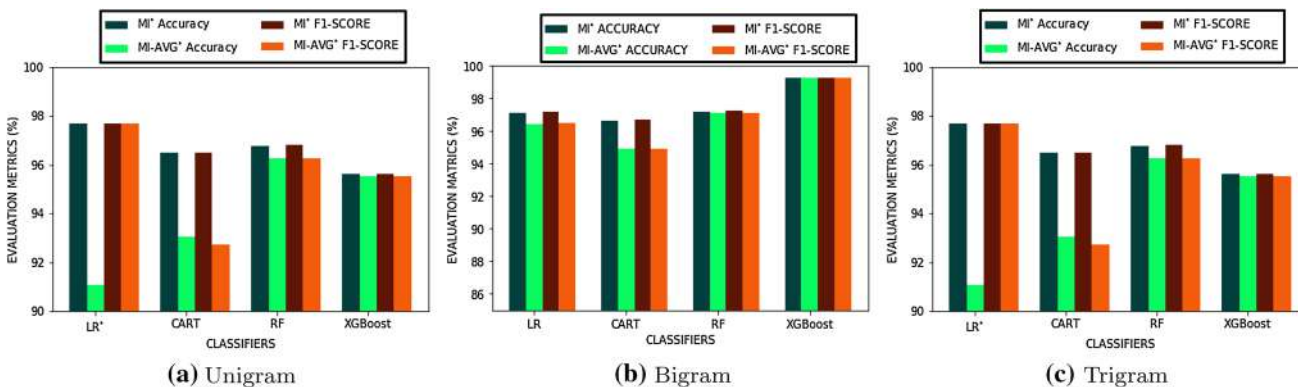


Fig. 6 Performance of MI* feature selection on unigram, bigram, and trigram

96.32%. In the case of bigram with XGBoost, accuracy and F1-score were 99.4% and 99.4% respectively.

Further, considering system call trigram with LR, the best accuracy achieved was 99.34% along with F1-score of 99.34%. Considering MI feature selector for unigram with Random Forest, the best accuracy attained was 96.75% and obtained F1-score of 96.79%. In the case of bigram, the best accuracy and F1-score obtained with XGBoost was 98% and 97.9% respectively. Whereas with LR, trigram

achieved accuracy and F1-score of 99.4% and 99.5% respectively.

Summary The results lead to the conclusion that the proposed feature selection method could derive attributes that had a higher correlation with the target class. Thus, resulting in improved outcome. In the case of unigram, Random Forest achieved higher accuracy and F1-score comparing other classifiers. Further for bigram, among all four classifiers, XGBoost exhibited better accuracy and F1-score. Comparing the results of trigram, LR achieved better

accuracy and F1-score compared to other classifiers. This is because, LR exhibited improved results as the feature vectors become linear in higher dimensional attribute space. The trend was clearly evident for the three Logistic regression model (i.e., unigram, bigram and trigram). Intuitively, in all cases of classification algorithms trained using bigrams and trigrams reported better performance compared to unigram.

5.4 Experiment IV: performance with DL classifier

In this experiment, DL was used to distinguish malware and legitimate applications. This experiment was conducted to analyze the performance of DNN compared to conventional ML approaches. A DNN model named DNN-2L with two hidden layers was designed. The first hidden layer consisted of 50% of attributes of the input feature space, as the number of neurons. Subsequently, the second layer contained 50% of the neurons, that were present in the previous layer. For example, if the feature set contained 1000 attributes, then the first layer will be created with 500 neurons and the second layer will be formed using 250 neurons. In DNN-2L, all layers contained Rectified Linear Unit (ReLU) activation function. Sigmoid activation function was used in the output layer, as malware identification is a binary classification problem. For faster convergence and to avoid overfitting, Adam optimization algorithm and cross entropy loss function was utilized respectively. Employing DNN-2L comprehensive analysis was conducted under following experimental settings and we investigated: (i) effect of dropout rate in the performance of model (Sect. 5.4.1) and (ii) effect of learning rate in the performance of model (Sect. 5.4.2).

5.4.1 Effect of dropout rate in the performance of model

To avoid overfitting on training data, dropout was proposed by Srivastava et al. [28]. Dropout is a regularization technique in which randomly selected neurons are removed during the training phase. This indicates that, the contribution of such neurons will be temporarily removed, during the forward pass and weight update will be ignored, in the backward pass. During the learning phase, a neuron specific to a particular layer relies on the neighbouring neurons. In a fully connected topology, a neuron tuned to specific feature was passed on to the upstream neurons. Thus, the network becomes more specific to the training data. On the contrary, if certain neurons were randomly eliminated, then the predictions are performed with the existing neurons. This suggests that many new patterns/representations, will be created and subsequently learned by the network. Thus, the network would be less sensitive

to the weights of neurons and less likely to overfit the training data.

To study the impact of dropout rate, a diverse classification model learned with attributes derived by our proposed feature selection method was created. In particular, the well known feature selectors were improved by deriving the call/sequence of calls having the ability to identify target classes. The outcome of the results are shown in Fig. 7. The results in Fig. 7a indicates that dropout rates of 0.2, 0.5 and 0.6 gave the best results for all three categories of features (i.e., unigram, bigram and trigram). An identical trend can be observed in Fig. 7b and c, i.e., better results are obtained at dropout rate of 0.3, 0.5 and 0.6. Finally, it was observed that DNN-2L learned with unigram, bigram and trigram at dropout rate of 0.2, 0.5 and 0.7 respectively, attained the best results as depicted in Fig. 7d. A similar observation has been observed when considering the F1-score as the evaluation metric, to evaluate the performance of the DL classifiers as show in Fig. 8.

5.4.2 Effect of learning rate in the performance of model

Further the importance of learning rate on the results of classification was explored. Learning rate is a hyper-parameter which denotes how much a model needs to be modified each time, by adjusting the weight. Lower value of learning rate indicates more time spent on training or in particular more steps needed to reach local minima. Conversely, large gradient descent learning rate would overshoot, besides missing local minima. Specifically, the model would fail to converge. In this study, the learning rate was varied such that it began with a small value (i.e. 0.001) and progressively increased by 0.01 until the maximum value of learning rate (i.e., 0.3) was reached (refer to Table 2).

5.5 Experiment V: evaluation against adversarial examples

The performance of any ML based system might degrade over time, eventually fortifying the system. In order to evaluate the detection capability of the classifier in the presence of adversarial samples, synthetic malware's mimicking statistical properties of legitimate set were created. In this context, adversarial malware samples were developed, by injecting varied proportion of prominent system calls, frequently invoked by the benign applications. The point of argument here is that, the classification algorithms fail to detect adversarial malware samples. The results show that as the proportion of system calls injected into each of the samples was increased, the recall declined dramatically, indicating that the classification algorithms fail to detect the malware samples (refer Table 3).

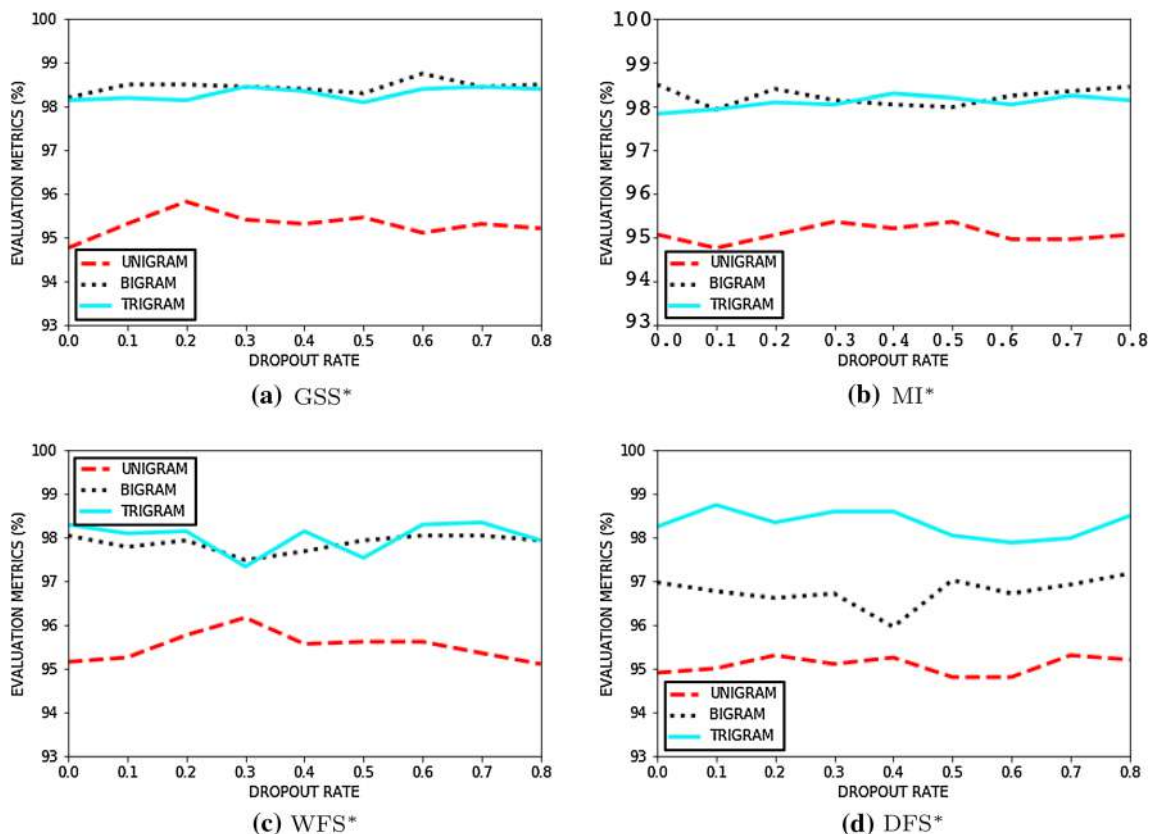


Fig. 7 Accuracy comparisons for different DL classifiers of proposed feature selection approach using dropout rate

In this experiment, adversarial samples for five detection models were generated. For each model, the recall rate was evaluated for 1%, 2% and 3% of prominent Benign system calls that were used to poison the malware samples. Figure 9a shows that the recall rate for Unigram in CART prior to the attack was 96.55% and after the attack, it got reduced to 33.73%. This was the case of 1% of prominent benign system calls that were appended to the malware samples for poisoning and the trend continued for both 2% and 3% of prominent benign system calls. Considering 2% of benign system calls, after poisoning the recall rate was minimized to 58.94% and for 3% of benign system calls it dropped to 30.89%. In case of a bigram, the recall rate reduced from 97.26 to 29.26% for 1% of benign system calls and for 2% benign system calls, the recall rate diminished to 30.0%. Likewise, for 3% of benign system calls, recall was minimized to 24.79%. Further for trigram, the recall rate for 1% of benign system calls reduced from 98.88 to 45.93%. Whereas for 2% of benign system calls, the rate minimized to 57.48% and for 3% of benign system calls, the rate reduced to 39.67 %.

Figure 9b depicts that for 1% benign system calls of unigram, LR shows a recall rate of 98.68% prior to poisoning, and after poisoning it was reduced to 50.4%. In case of 2% and 3% of system calls, the true positive rate

was reduced to 68.29% and 66.66% respectively. For bigram after poisoning, the recall rate was reduced from 98.78 to 77.64%, 81.3% and 69.1% for 1%, 2% and 3% of benign system calls. Considering trigram, from 99.79% of recall rate, it diminished to 86.17%, 86.6% and 85.02% for 1% 2% and 3% of prominent benign system calls respectively.

Figure 9c shows that in case of Random Forest, for unigram the true positive rate declined from 97.89 to 61.38%, 74.79% and 35.77% for 1%, 2% and 3% of benign system calls respectively. Considering 1%, 2% and 3% of benign system calls of bigram, the recall rate reduced from 98.38 to 79.26%, 83.33%, and 74.39% respectively. Further, for poisoned trigram, the recall rate diminished from 98.58 to 82.11%, 84.61% and 84.61% for 1%, 2% and 3% of benign system calls.

From Fig. 9d it is clear that with XGBoost, the recall rate of unigram reduced from 95.8 to 80.01% for 1%, 2% and 3% of benign system calls after poisoning. Considering 1%, 2% and 3% of benign system calls in Bigram after poisoning, the recall rate reduced from 98.7 to 92.22%. In case of Trigram, for 1%, 2% and 3% of benign system calls, the recall rate of 99% diminished to 91.9%.

Figure 10 depicts that for unigram, the recall rate diminished from 95.2 to 44.53% in case of 1% of

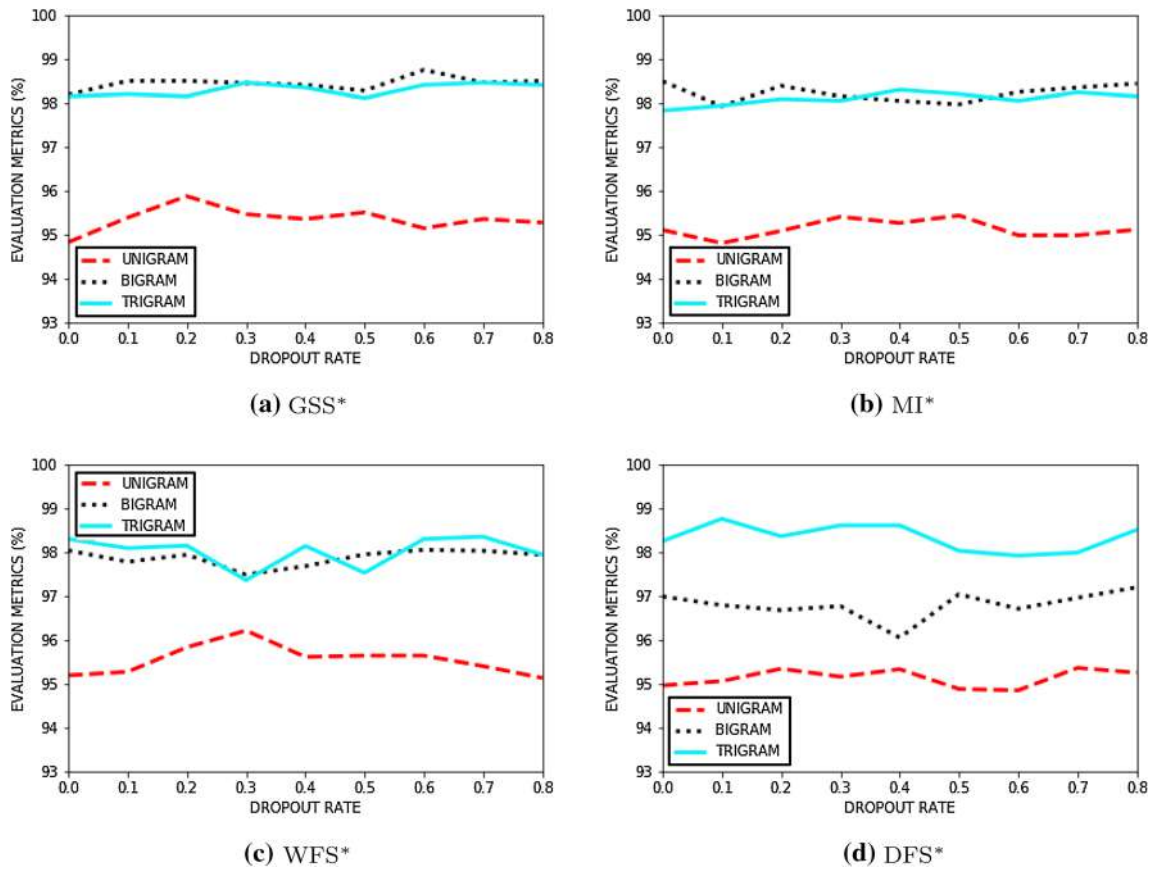


Fig. 8 F1-Score comparisons for different DL classifiers of proposed feature selection approach using dropout rate

Table 2 Performance of DL classifier with proposed feature selection approaches using dropout rate and learning rate

Feature selectors	Category	Drop out	LR	\mathcal{A}	\mathcal{F}_1	\mathcal{PRC}	\mathcal{REC}
GSS*	UNIGRAM	0.2	0.1	0.952	0.952	0.950	0.956
	BIGRAM	0.6	0.2	0.990	0.990	0.990	0.994
	TRIGRAM	0.8	0.1	0.980	0.980	0.980	0.990
DFS*	UNIGRAM	0.7	0.001	0.950	0.950	0.953	0.943
	BIGRAM	0.8	0.1	0.990	0.990	0.984	0.992
	TRIGRAM	0.1	0.1	0.980	0.981	0.970	0.991
MI*	UNIGRAM	0.5	0.001	0.952	0.952	0.952	0.952
	BIGRAM	0.8	0.1	0.980	0.980	0.980	0.982
	TRIGRAM	0.4	0.001	0.987	0.987	0.986	0.989

Asterisks indicate a revised feature set after the application of SAILS

prominent benign system calls and for 2% and 3% of benign system calls, the recall rate reduced to 43.3% and 42.1%. Considering 1% of injected bigram, the recall rate minimized from 98.2 to 40.48%, the recall at 2% and 3% injection rate reduced to 34.81% and 32.39% respectively. Further for trigram, in case of 1% injected calls, the recall declined from 98.58 to 48.18%, moreover, for 2% to 3%, the recall dropped from 42.51 and 37.65% respectively.

6 Conclusion and future directions

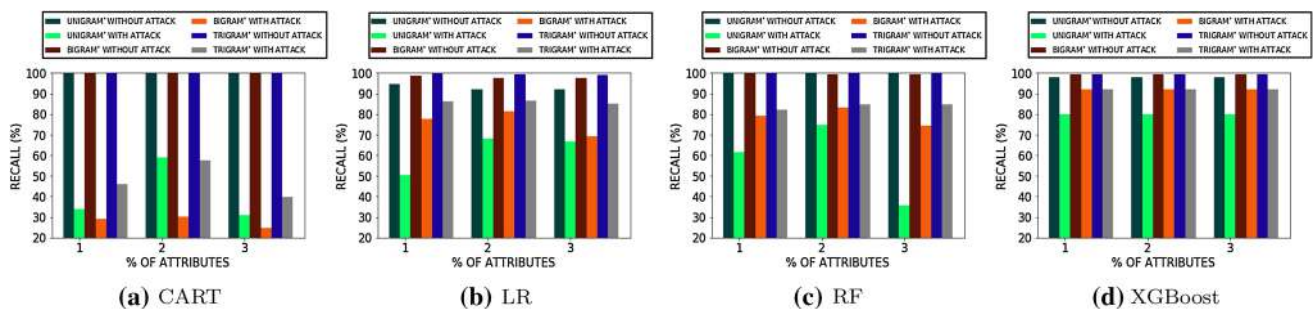
In this paper, a new feature selection method, SAILS, is designed, which can provide better results compared to conventional feature selection approaches. Also, the classifier performance of different N -grams is studied. We performed parallel analysis of the Android malware detector using deep learning network and machine learning algorithms. Further, the performance of the model was evaluated by analysing the change in dropout and learning rate.

Table 3 Performance comparison of WFS feature selection method with other approaches

Classifier	Feature selectors	Unigram		Bigram		Trigram	
		FL	\mathcal{A}	FL	\mathcal{A}	FL	\mathcal{A}
LR	WFS*	40(37)	89.3	100(2212)	98.0	100(20354)	99.4
	MI-AVG	100(98)	91.05	100(2212)	96.41	90(18318)	96.75
	GSS-AVG	90(88)	88.43	100(2212)	99.08	100(20354)	99.34
	DFS-AVG	100(98)	91.96	90(1990)	97.16	100(20354)	99.34
CART	WFS*	40(37)	93.1	100(2212)	96.7	100(20354)	98.7
	MI-AVG	100(98)	92.26	100(2212)	94.94	100(20354)	94.78
	GSS-AVG	90(88)	92.57	80(1769)	95.08	100(20354)	98.17
	DFS-AVG	90(88)	92.02	90(1990)	95.39	90(18318)	97.89
RF	WFS*	90(88)	97.2	20(442)	97.6	100(20354)	97.5
	MI-AVG	100(98)	96.1	80(1769)	97.21	100(20354)	96.8
	GSS-AVG	100(98)	95.5	100(2212)	98.53	100(20354)	97.77
	DFS-AVG	100(98)	96.11	100(2212)	97.21	100(20354)	97.97
XGBoost	WFS*	70(68)	96.6	40(884)	97.9	20(4070)	97.7
	MI-AVG	100(98)	95.5	90(1990)	99.3	100(20354)	99.2
	GSS-AVG	80(78)	92.65	70(1548)	97.67	80(16283)	97.72
	DFS-AVG	90(88)	95.4	100(2212)	97.3	80(16283)	98.2

FL denotes feature length at which the best outcomes were obtained. FL is expressed in the form of $P(Q)$, where P denotes the percentage of features extracted from the feature space and Q denote the number of attributes used to create the model

Asterisks indicate a revised feature set after the application of SAILS

**Fig. 9** Performance evaluation of unigram, bigram and trigram after poisoning using ML

Adversarial attacks are also performed on the ML models. It is observed that the adversary could deceive the current ML based malware detectors. A drop in performance is observed when the trained models were given evasive examples as input. Thus, it is important to develop robust ML models trained with adversarial patterns, such that Android malware detectors are capable of recognizing tainted samples.

Our current work focuses on dynamic analysis of Android malware. In future work, we envisage the use hybrid analysis on a larger dataset. It is also planned to

include features relating to network packets (packet size, packet payload size, packet inter-arrival time, TCP flag status, the total number of bytes in packets, packet direction, protocols, etc.), to train ML algorithms. The collection of these features along-with systems calls would undoubtedly reveal promising patterns for identifying malware. We also plan to model new feature selection techniques having high correlation with class, but loosely correlated with other features. Finally, we plan to carry out attacks on classifier ensembles, and develop

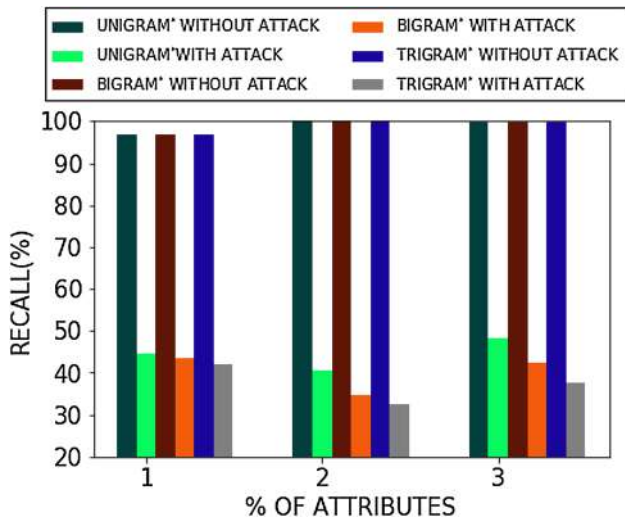


Fig. 10 Performance evaluation of unigram, bigram and trigram after poisoning using DL

countermeasures to harden classifier for minimizing misclassification rate.

Compliance with ethical standards

Conflicts of interest There is no conflict of interest for the paper.

Appendix

In this section, different scores of N-gram for GSS, WFS, and MI are presented to illustrate the malware and benign samples for various features (see the Figs. 11, 12 and 13).

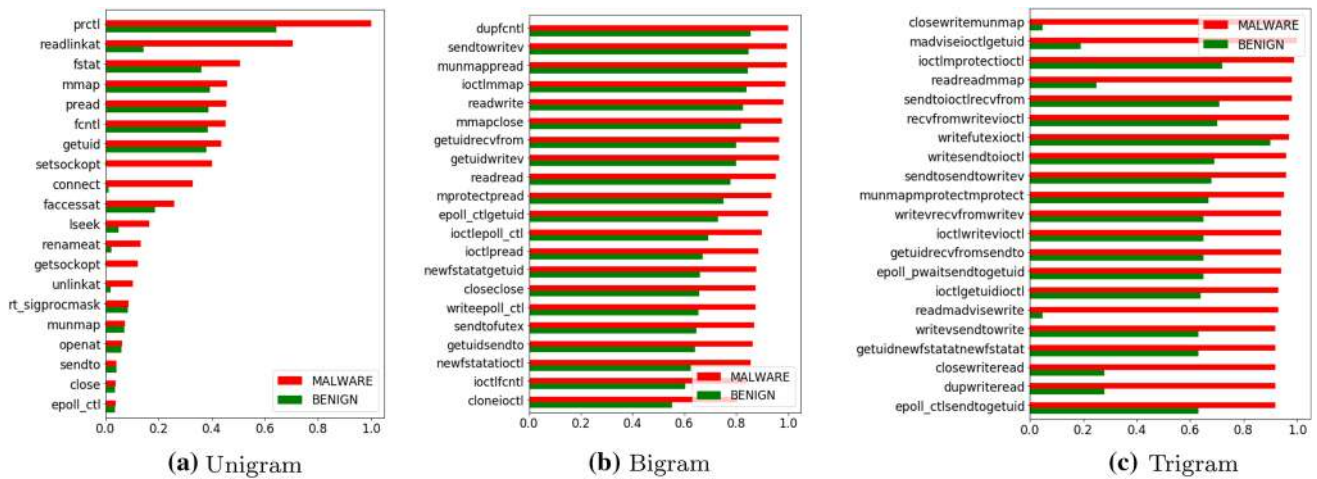


Fig. 11 Score difference of N-grams for GSS*

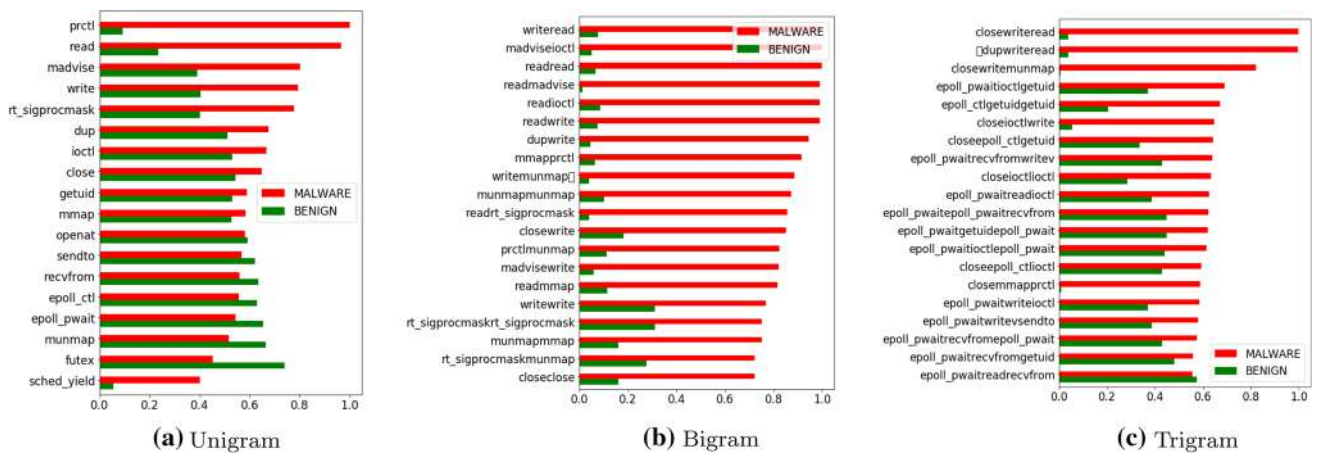


Fig. 12 Score difference of N-grams for WFS*

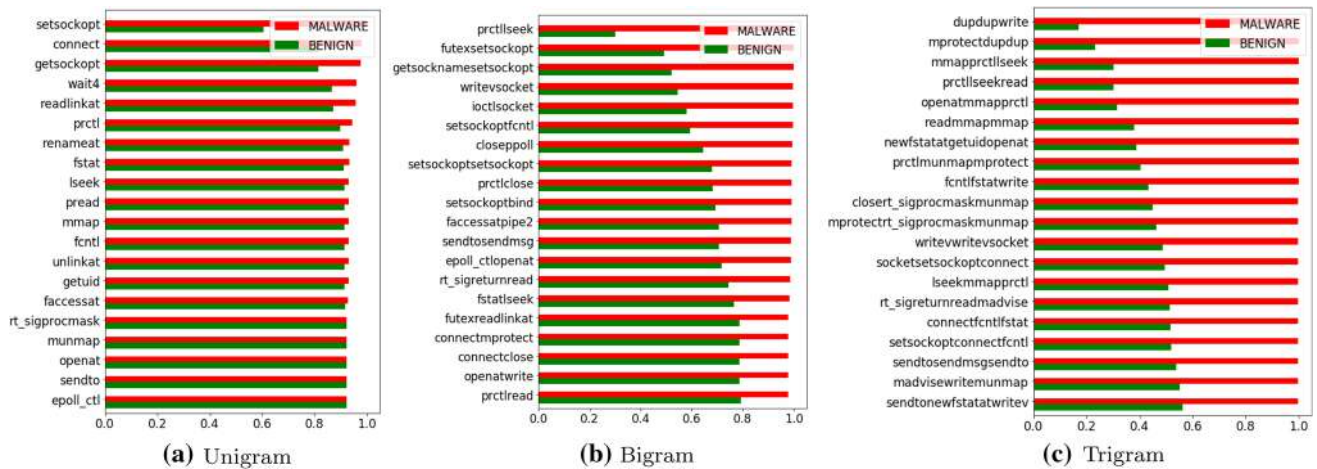


Fig. 13 Score difference of N-grams for MI*

References

1. Aafer, Y., Du, W., Yin, H.: Droidapiminer: mining api-level features for robust malware detection in android. In: International Conference on Security and Privacy in Communication Systems, pp. 86–103. Springer, Berlin (2013)
2. Afonso, V.M., de Amorim, M.F., Grégio, A.R.A., Junquera, G.B., de Geus, P.L.: Identifying android malware using dynamically obtained features. *J. Comput. Virol. Hacking Tech.* **11**(1), 9–17 (2015)
3. Arp, D., Spreitzenbarth, M., Hubner, M., Gascon, H., Rieck, K., Siemens, C.: Drebin: effective and explainable detection of android malware in your pocket. *Ndss* **14**, 23–26 (2014)
4. Arshad, S., Shah, M.A., Wahid, A., Mehmood, A., Song, H., Yu, H.: Samadroid: a novel 3-level hybrid malware detection model for android operating system. *IEEE Access* **6**, 4321–4339 (2018)
5. Bhandari, S., Panihar, R., Naval, S., Laxmi, V., Zemmari, A., Gaur, M.S.: Sword: semantic aware android malware detector. *J. Inf. Secur. Appl.* **42**, 46–56 (2018)
6. Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., Giacinto, G., Roli, F.: Evasion attacks against machine learning at test time. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 387–402. Springer, Berlin (2013)
7. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
8. Burguera, I., Zurutuza, U., Nadjm-Tehrani, S.: Crowddroid: behavior-based malware detection system for android. In: Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices, pp. 15–26. ACM (2011)
9. Cyber security facts and statistics for 2019. <https://us.norton.com/internetsecurity-emerging-threats-10-facts-about-todays-cybersecurity-landscape-that-you-should-know.html> (2019). Accessed 10 Aug 2019
10. Cao, Y., Yang, J.: Towards making systems forget with machine unlearning. In: 2015 IEEE Symposium on Security and Privacy, pp. 463–480. IEEE (2015)
11. Chen, L., Hou, S., Ye, Y., Chen, L.: An adversarial machine learning model against android malware evasion attacks. In: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data, pp. 43–55. Springer, Berlin (2017)
12. Chen, S., Xue, M., Fan, L., Hao, S., Xu, L., Zhu, H., Li, B.: Automated poisoning attacks and defenses in malware detection systems: an adversarial machine learning approach. *Comput. Secur.* **73**, 326–344 (2018)
13. Chen, T., Guestrin, C.: Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)
14. Grosse, K., Papernot, N., Manoharan, P., Backes, M., McDaniel, P.: Adversarial examples for malware detection. In: European Symposium on Research in Computer Security, pp. 62–79. Springer, Berlin (2017)
15. Han, W., Xue, J., Wang, Y., Huang, L., Kong, Z., Mao, L.: Maldae: detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. *Comput. Secur.* **83**, 208–233 (2019)
16. Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression, vol. 398. Wiley, New York (2013)
17. Hou, S., Saas, A., Chen, L., Ye, Y.: Deep4maldroid: a deep learning framework for android malware detection based on linux kernel system call graphs. In: 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), pp. 104–111. IEEE (2016)
18. Hou, S., Saas, A., Ye, Y., Chen, L.: Droiddelver: an android malware detection system using deep belief network based on API call blocks. In: International Conference on Web-Age Information Management, pp. 54–66. Springer, Berlin (2016)
19. Hou, S., Ye, Y., Song, Y., Abdulhayoglu, M.: Hindroid: an intelligent android malware detection system based on structured heterogeneous information network. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1507–1515. ACM (2017)
20. Ishibashi, H., Hihara, S., Iriki, A.: Acquisition and development of monkey tool-use: behavioral and kinematic analyses. *Can. J. Physiol. Pharmacol.* **78**(11), 958–966 (2000)
21. Largeton, C., Moulin, C., Géry, M.: Entropy based feature selection for text categorization. In: Proceedings of the 2011 ACM Symposium on Applied Computing, pp. 924–928. ACM (2011)
22. Mobile malware evolution 2019. <https://securelist.com/mobile-malware-evolution-2018/89689/> (2019). Accessed 10 Aug 2019
23. Michael, S., Florian, E., Thomas, S., Felix, C.F., Hoffmann, J.: Mobilesandbox: looking deeper into android applications. In: Proceedings of the 28th International ACM Symposium on Applied Computing (SAC) (2013)
24. Naway, A., Li, Y.: A review on the use of deep learning in android malware detection. arXiv preprint arXiv:1812.10360 (2018)

25. Roundy, K.A., Miller, B.P.: Hybrid analysis and control of malware. In: International Workshop on Recent Advances in Intrusion Detection, pp. 317–338. Springer, Berlin (2010)
26. Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **21**(3), 660–674 (1991)
27. Santos, I., Peña, Y.K., Devesa, J., Bringas, P.G.: N-grams-based file signatures for malware detection. *ICEIS* **2**(9), 317–320 (2009)
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
29. Suci, O., Marginean, R., Kaya, Y., Daume III, H., Dumitras, T.: When does machine learning FAIL? Generalized transferability for evasion and poisoning attacks. In: 27th USENIX Security Symposium (USENIX Security 18), pp. 1299–1316 (2018)
30. Tong, F., Yan, Z.: A hybrid approach of mobile malware detection in android. *J. Parallel Distrib. Comput.* **103**, 22–31 (2017)
31. VirusTotal. <http://virustotal.com/> (2019). Accessed 10 Aug 2019
32. Wang, W., Zhao, M., Wang, J.: Effective android malware detection with a hybrid model based on deep autoencoder and convolutional neural network. *J. Ambient Intell. Humaniz. Comput.* **10**(8), 3035–3043 (2019)
33. Yang, Y., Shen, H.T., Ma, Z., Huang, Z., Zhou, X.: L2, 1-norm regularized discriminative feature selection for unsupervised. In: Twenty-Second International Joint Conference on Artificial Intelligence (2011)
34. Yuan, Z., Lu, Y., Xue, Y.: Droiddetector: android malware characterization and detection using deep learning. *Tsinghua Sci. Technol.* **21**(1), 114–123 (2016)
35. Zhang, J., Zhang, K., Qin, Z., Yin, H., Wu, Q.: Sensitive system calls based packed malware variants detection using principal component initialized multilayers neural networks. *Cybersecurity* **1**(1), 10 (2018)
36. Zheng, Z., Wu, X., Srihari, R.: Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explor. Newslett.* **6**(1), 80–89 (2004)
37. Zhou, Y., Jiang, X.: Dissecting android malware: characterization and evolution. In: 2012 IEEE Symposium on Security and Privacy, pp. 95–109. IEEE (2012)
38. 9apps: Android app website. <https://www.9apps.com/> (2019). Accessed 10 Aug 2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ananya Asok graduated from SCMS School of Engineering and Technology, Kochi, Kerala in May 2019 with a Bachelor's of Technology in Computer Science and Engineering. She has keen interests in the area of machine learning. Her main research interest is Android Mobile security and privacy and machine learning.



Aswathy A. graduated from SCMS School of Engineering & Technology in July 2019 with a bachelor's of technology in Computer Science and Engineering. She has a keen interest in the field of machine learning and her involvement in the developer community ecosystem had made a great impact on the lives of many aspiring women engineers. Her main research interest is Android Mobile security and privacy and machine learning.



Amal T.R. received his Bachelor of Science in Mathematics from Calicut University in 2016. Also, he received a Master's degree in Computer Applications from APJ Abdul Kalam Technical University in 2019. His main research interest is Android Mobile security and privacy and machine learning.



Swathy P.G. graduated from SCMS School of Engineering and Technology in July 2019 with a Master's in Computer Science and Information Systems. Her main research interest is Android Mobile security and privacy and machine learning.



Vinod P. is Professor in Department of Computer Science & Engineering at SCMS School of Engineering and Technology, Kerala, India. He was Post Doc at Department of Mathematics, University of Padua, Italy. During his Post Doctoral research he was actively involved in executing EU-H2020 project TagitSmart. He holds his Ph.D in Computer Engineering from Malaviya National Institute of Technology, Jaipur, India. He has more

than numerous research articles published in peer reviewed Journals and International Conferences. He is reviewer of number of security journals, and has also served as programme committee member in the

International Conferences related to Computer and Information Security. His current research is involved in the development of malware scanner for mobile application using machine learning techniques. Vinod's area of interest is Adversarial machine learning, Malware analysis, context aware privacy persevering data mining, ethical hacking and natural language processing.



Mohammad Shojarf is an Intel Innovator, a Senior IEEE member, a Senior Researcher and a Marie Curie Fellow in the SPRITZ Security and Privacy Research group at the University of Padua, Italy. Also, he was CNIT Senior Researcher at the University of Rome Tor Vergata contributed to 5G PPP European H2020 “SUPERFLUIDITY” project. He is a PI on PRISE-NODE project, a 275,000 euro Horizon 2020 Marie Curie project in the areas of network

security and Fog computing and resource scheduling collaborating

between the University of Padua and University of Melbourne. He also was a PI on an Italian SDN security and privacy (60,000 euro) supported by the University of Padua in 2018. He also contributed to some Italian projects in telecommunications like GAUChO—A Green Adaptive Fog Computing and Networking Architecture (400,000 euro), and SAMM-Clouds- Secure and Adaptive Management of Multi-Clouds (30,000 euro) collaborating among Italian universities. He received a Ph.D. in ICT from Sapienza University of Rome, Italy, in 2016 with an “Excellent” degree. His main research interests are in the area of Computer Networks, Network Security, and Privacy. In this area, he published more than 100+ papers in topmost international peer-reviewed journals and conferences, e.g., IEEE TCC, IEEE TNSM, IEEE TGCN, IEEE TSUSC, IEEE Network, IEEE SMC, IEEE PIMRC, and IEEE ICC/GLOBECOM. He served as a PC member of several prestigious conferences, including IEEE INFOCOM Workshops in 2019, IEEE GLOBECOM, IEEE ICC, IEEE UCC, IEEE ScalCom, and IEEE SMC. He was GC in FMEC 2019, INCoS 2019, INCoS 2018, and a Technical Program Chair in IEEE FMEC 2020. He served as an Associate Editor in IEEE Transactions on Consumer Electronics, IET Communication, Springer Cluster Computing, and Ad Hoc & Sensor Wireless Networks Journals.

Received December 3, 2020, accepted December 18, 2020, date of publication December 24, 2020, date of current version January 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047136

Secrecy Outage Probability of Relay Selection Based Cooperative NOMA for IoT Networks

HUI LI¹, YAPING CHEN¹, MINGFU ZHU², JIANGFENG SUN³, (Member, IEEE),
DINH-THUAN DO⁴, VARUN G. MENON⁵, (Senior Member, IEEE),
AND SHYNU P. G.⁶, (Member, IEEE)

¹School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China

²Huawei-Chuaitian 5G Edge Computing Laboratory, Hebei 458000, China

³College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454003, China

⁴Department of Computer Science and Information Engineering, College of Information and Electrical Engineering, Asia University, Taichung 41354, Taiwan

⁵Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India

⁶School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India

Corresponding author: Jiangfeng Sun (sunjiangfeng@bupt.edu.cn)

This work was supported in part by the Basic and Advanced Technology Research Project of Henan Province under grant 152300410103, in part by Science and Technology Research Project of Henan Province under grant 202102310299, and in part by Opening Project of Henan Engineering Laboratory of Photoelectric Sensor and Intelligent Measurement and Control of Henan Polytechnic University under grant HELPSIMC-2020-002.

ABSTRACT As an important partner of fifth generation (5G) communication, the internet of things (IoT) is widely used in many fields with its characteristics of massive terminals, intelligent processing, and remote control. In this paper, we analyze security performance for the cooperative non-orthogonal multiple access (NOMA) networks for IoT, where the multi-relay Wyner model with direct link between the base station and the eavesdropper is considered. In particular, secrecy outage probability (SOP) for two kinds of relay selection (RS) schemes (i.e., single-phase RS (SRS) and two-phase RS (TRS)) is developed in the form of closed solution. As a benchmark for comparison, the SOP for random RS (RRS) is also obtained. To gain more meaningful insights, approximate derivations of SOP under the high signal-to-noise ratio (SNR) region are provided. Results of statistical simulation confirm the theoretical analysis and testify that: i) Compared with RRS scheme, SRS and TRS may improve secure performance because of obtaining smaller SOPs; ii) There exists secrecy performance floor for the SOP in strong SNR regime, which is dominated by NOMA protocol; iii) The security performance can be enhanced by augmenting the quantity of relays for SRS and TRS strategies. The purpose of this work is to provide theoretical basis for the analysis and design of anti-eavesdropping for NOMA systems in IoT.

INDEX TERMS Non-orthogonal multiple access, physical layer security, secrecy outage probability, single-phase relay selection, two-phase relay selection.

I. INTRODUCTION

Recently, the rapid development of IoT makes all walks of life get convenient and fast services. However, due to the importance of ownership and privacy protection, the IoT system must provide corresponding security mechanisms. The classical method to deal with the security problem is complex encryption and decryption scheme [1]. Quantum computing can crack complex keys. Moreover, the terminals of IoT are often limited in size and power, and do not have strong computing power. These contradictions lead that the classical method is not so effective in many scenarios [2]. So an alternative mechanism, i.e., physical layer security (PLS) exhibits

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenyu Zhou¹.

more advances. The method of PLS was initially discussed by Wyner from the standpoint of information theory [3]. PLS is a new approach to enhance network security by utilizing the characteristics of channels, which has caught quantity of attention due to the randomness of fading channels rather than encryption technology [4]–[7]. In [8], the authors studied the secrecy behaviours for underlay cooperative relaying networks. Recently, NOMA is deemed to have a bright prospect in 5G networks on account that it can improve the band-efficient and spectral efficiency [9]–[13]. Serving multiple users working at the same frequency band with different power-split is the core thought of NOMA [14]. Do *et al.* put forward a model which can serve cellular networks better in NOMA [15]. The authors of [16] discussed a large-scale network with an antenna and multiple antennas in NOMA

systems, and derived the SOP. Lei *et al.* researched a security NOMA system including two different forms of eavesdropping [17]. The ambient backscatter NOMA systems was studied in terms of the secure performance [18]. Jiang *et al.* analyzed the secure performance for uplink NOMA including multiple eavesdroppers in [19]. Therefore, exploring PLS in NOMA systems has also aroused the interests of many researchers.

Cooperative communication is a specially efficient method by furnishing greater diversity and expanding network coverage [20]. At present, two fields are mainly included in cooperative communication for NOMA's research. On the one hand, the use of NOMA in cooperative networks was discussed in [21]–[24]. On the other hand, cooperative NOMA was first put forward by Ding *et al.* in [25] and researched in [26]–[29]. Choi studied the transmission rates on the cooperative system in [21]. The authors studied the interruption probability (IP) and systematic capacity of NOMA using decoded and forward (DF) in relaying systems [22]. In [23], the SOP was investigated based full-duplex (FD) in cooperative communication using optimizing power allocation jointly. Men *et al.* discussed the outage character using amplify and forward (AF) protocol on Nakagami- m distribution for NOMA in [24]. The core idea of cooperative NOMA is that nearby NOMA users are treated as DF relaying to transfer the messages for far NOMA users. The secrecy behaviors for both AF and DF relaying strategies were investigated in cooperative NOMA system [26]. Simultaneous wireless information and power transmission (SWIPT) was adopted by nearby NOMA users which were counted as DF relaying [27]. The work researched the security transmission and proposed an optimal power distribution scheme with maximum secrecy sum rate, where the precondition was that the users' quality of service (QoS) met the conditions [28]. The authors of [29] employed FD and artificial noise (AN) methods in two-way relaying networks based on NOMA. The mathematical expressions for the ergodic secrecy rate were discussed under containing and excluding eavesdroppers.

As a popular transmission scheme, relay selection (RS) has the advantages of low complexity due to taking full advantage of spatial variety and high spectrum-efficient [30], [31]. Considering that there might be some differences between two users in the QoS requirements, two-stage single-relay-selection and dual-relay-selection strategies were put forward, respectively [32]. Ding *et al.* derived closed-form expressions for the precise and asymptotic outage probability (OP) by employing single-stage RS and two-stage RS strategies in cooperative NOMA. And the two NOMA users were classified as nearby and far users by their QoS, rather than their channel conditions [33]. Accurate analytical formulae for the OP and IP were analyzed by using two relay selection strategies (i.e., optimal RS and suboptimal RS) in wireless communication networks (WCNs) [34]. Under three wiretapping cases including one eavesdropper, non-colluding and colluding eavesdroppers, the secrecy outage behaviors of the TRS strategy based on the system over Nakagami- m

fading channels were investigated in [35]. Zhang *et al.* analyzed the SOP with optimal relay selection, suboptimal relay selection and multiple relays uniting schemes. In addition, the confidentiality of a cognitive DF relaying network over Nakagami- m fading channels with independent but not necessarily identical distributed was also surveyed in [36].

Although these previous contributions provided a firm foundation for understanding collaborative NOMA and RS technologies, it still needs further developments and applications. It should be pointed out that RS schemes can meet the requirements of actual IoT situation. In this paper, we investigate the SRS and TRS methods which can achieve the minimum SOP. As far as we know, there is no research on the security performance of SRS and TRS schemes in cooperative NOMA networks considering direct link between base station and eavesdropper. To this end, we explore SOP using RS schemes for basing on half-duplex (HD) NOMA networks over independently Rayleigh distribution. More specifically, the rate of data transmission for the far user is assured to choose a relay as its auxiliary equipment to forward the messages in the SRS strategy. Under the premise of guaranteeing the data transmission rate of far user, the maximum data rate of the service is provided for nearby user to select the relay opportunistically in the TRS scheme. The key contributions of this paper are summarized as follows:

- This paper describes system model of cooperative NOMA for IoT and focuses on two kinds of relay selection strategies (i.e., SRS and TRS schemes). Moreover, the direct link between the eavesdropper and the base station is considered. The eavesdropper uses selective combination (SC) technique to process the received signals from two slots.
- The theoretical SOP is derived by employing the SRS strategy over Rayleigh fading channel. In addition, the SOP for RRS scheme is also analyzed as a contrast. The results show that SRS strategy obtains the lower SOP. To better understand secure outage performance, the asymptotic behaviors of SOP are analyzed with RRS and SRS schemes in cooperative NOMA.
- We also derive the formulas of SOP for TRS scheme in cooperative NOMA based on HD. What's more, experimental results prove that TRS scheme can obtain the superior SOP. To get more insights, the approximate SOP of TRS scheme under high SNR regime is analyzed in cooperative NOMA. The results also verify that the security performance can be enhanced distinctly by augmenting the quantity of relays.

The specific arrangement of each section is as follows. In Section II, the network system of HD NOMA's RS schemes is established. Section III deduces new analytic formulae of SOP for the RRS, SRS and TRS schemes. In Section IV, the asymptotical SOPs in high SNR regime are derived. Section V presents numerical results and systematic performance. The conclusions are shown in Section VI in the paper.

Notations: The $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex Gaussian distribution with expectation μ and standard variance σ .

The $\Pr(\cdot)$ and $\mathbb{E}(\cdot)$ are the probability and expectation operation. $f_X(\cdot)$ and $F_X(\cdot)$ are the probability density function (PDF) and the cumulative distribution function (CDF), respectively.

II. SYSTEM MODEL

As illustrated in Fig.1, a typical dual-hop NOMA relaying system for IoT includes a base station (*BS*), K half-duplex relays, a couple of legitimate users (e.g., the nearby user D_1 and far user D_2) and one eavesdropper (*Eve*). It should be noted that the direct links from *BS* to D_j ($j = 1, 2$) are not considered, but the direct link between *BS* and *Eve* exists. So, the *Eve* processes the received signals by using SC arithmetic. Adopting a multi-access scheme, multiple users can be easily partitioned into many groups in this cooperative model, each of these groups implements the NOMA protocol [37]. In the network model, all relaying nodes are equipped with receiving and transmitting antennas, but *BS* and users have only one antenna. The *BS* tries to communicate the users via relays, but there exists eavesdropping between transmissions, and the information leakage exists in the transmission between two slots. Each relay is assumed to use DF protocol. All wireless channels are affected by additive white Gaussian noise (AWGN) and modeled as independent non-selective Rayleigh distribution. The distance from X to Y is represented as d_{XY} , α denotes exponent for the path loss, h_{XY} denotes the channel coefficient from X to Y , $XY \in \{SR_i, R_iD_1, R_iD_2, SE, R_iE\}$ and $h_{XY} \sim \mathcal{CN}(0, 1)$. The PDF and CDF for $|h_{XY}|^2$ have an exponential distribution as

$$f_{|h_{XY}|^2}(x) = \frac{1}{g_{XY}} \exp\left(-\frac{x}{g_{XY}}\right), \quad (1)$$

and

$$F_{|h_{XY}|^2}(x) = 1 - \exp\left(-\frac{x}{g_{XY}}\right), \quad (2)$$

respectively, where g_{XY} is the mean channel power gain [38]. The two legitimate users are segmented into nearby and far users on the basis of their QoS. More specifically, with the assistance of relay chosen, the QoS requirements of legal users can be effectively provided for the IoT scenario. Therefore, we assume that D_1 can serve opportunely with low target data rates, D_2 needs to be served quickly.

During the first stage, the *BS* transmits composite messages $\sqrt{a_1 P_s} x_1 + \sqrt{a_2 P_s} x_2$ to the assistances on the basis of NOMA theory, and normalization method of x_1 and x_2 signal is adopted respectively, i.e. $\mathbb{E}(|x_1|^2) = \mathbb{E}(|x_2|^2) = 1$, P_s and P_r denote the transmitted power from the *BS* and R_i . a_1 and a_2 are the corresponding power allocation coefficients. In fact, in order to provide better fairness and QoS requirements among users [39], we hypothesize that $a_2 > a_1$ and $a_1 + a_2 = 1$. Hence the received messages at the i th relay R_i can be expressed as

$$y_{SR_i} = \frac{h_{SR_i}}{\sqrt{d_{SR}^\alpha}} \left(\sqrt{a_1 P_s} x_1 + \sqrt{a_2 P_s} x_2 \right) + n_{SR_i}, \quad (3)$$

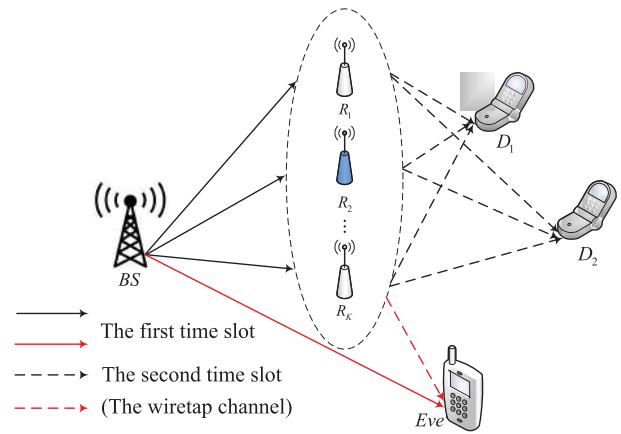


FIGURE 1. System model for IoT.

where n_{SR_i} is written as the superimposed Gaussian noise at relay i .

In order to reduce the interference for decoding signal x_1 of D_1 at R_i , the successive interference cancellation (SIC) method is employed to detect the information x_2 of D_2 firstly with the high power allocation coefficient. Therefore, the received signal-to-interference-plus-noise ratios (SINRs) to decode x_1 and x_2 at R_i are shown by

$$\gamma_{R_i, D_2} = \frac{a_2 \rho_s |h_{SR_i}|^2}{a_1 \rho_s |h_{SR_i}|^2 + d_{SR}^\alpha}, \quad (4)$$

and

$$\gamma_{R_i, D_1} = \frac{a_1 \rho_s |h_{SR_i}|^2}{d_{SR}^\alpha}, \quad (5)$$

where $\rho_x = \frac{P_x}{N_0}$, $x \in (s, r)$ is the transmit SNR, and N_0 is the mean power of the AWGN in this system.

In the same way, the message received at *Eve* can be expressed as

$$y_{SE} = \frac{h_{SE}}{\sqrt{d_{SE}^\alpha}} \left(\sqrt{a_1 P_s} x_1 + \sqrt{a_2 P_s} x_2 \right) + n_{SE}. \quad (6)$$

where n_{D_j} , n_X denote the Gaussian noise at users D_j and X ($X = SE, RE$).

We analyse the SINR for wiretapper to decode x_1 and x_2 . Considering a direct link between *BS* and *Eve* in this paper. Therefore, in this time slot, the instantaneous SINRs at *Eve* that eavesdrops the messages from legal users D_j are written by

$$\gamma_{SE_1} = \frac{a_1 \rho_s |h_{SE}|^2}{d_{SE}^\alpha}, \quad (7)$$

and

$$\gamma_{SE_2} = \frac{a_2 \rho_s |h_{SE}|^2}{a_1 \rho_s |h_{SE}|^2 + d_{SE}^\alpha}. \quad (8)$$

During the second stage, it is assumed that relay R_i can decode received messages and transmit signals to the target nodes, the following situations are met in this phase,

i) $\log\left(\frac{1+\gamma_{R_i,D_1}}{1+\gamma_{E_1}}\right) \geq R_{D_1}$, ii) $\log\left(\frac{1+\gamma_{R_i,D_2}}{1+\gamma_{E_2}}\right) \geq R_{D_2}$, where γ_{E_j} is the SNR at *Eve* and will be further analyzed later, R_{D_j} denotes the target data transmission rate. Selective relay R_i forwards the signals to the user, so the signals received in D_j can be represented as

$$y_{D_j} = \frac{h_j}{\sqrt{d_{RD_j}^\alpha}} \left(\sqrt{a_1 P_r x_1} + \sqrt{a_2 P_r x_2} \right) + n_{D_j}. \quad (9)$$

The signals received in *Eve* in this phase can be represented as

$$y_{RE} = \frac{h_{RE}}{\sqrt{d_{RE}^\alpha}} \left(\sqrt{a_1 P_r x_1} + \sqrt{a_2 P_r x_2} \right) + n_{RE}. \quad (10)$$

It is assumed that perfect SIC can be used in D_2 to detect messages from D_1 with a higher transmitting power. Therefore, D_2 detects the SINR of x_1 given by the following formula,

$$\gamma_{D_1,D_2} = \frac{a_2 \rho_r |h_1|^2}{a_1 \rho_r |h_1|^2 + d_{RD_1}^\alpha}. \quad (11)$$

Then, the received SINR at D_1 is given by

$$\gamma_{D_1} = \frac{a_1 \rho_r |h_1|^2}{d_{RD_1}^\alpha}. \quad (12)$$

Meanwhile, D_2 decodes messages x_2 by regarding x_1 as interference, and the SINR can be shown as

$$\gamma_{D_2} = \frac{a_2 \rho_r |h_2|^2}{a_1 \rho_r |h_2|^2 + d_{RD_2}^\alpha}. \quad (13)$$

For the second time slot, the instantaneous SINRs at *Eve* to wiretap the messages are expressed as

$$\gamma_{R_i E_1} = \frac{a_1 \rho_r |h_{R_i E}|^2}{d_{RE}^\alpha} \quad (14)$$

$$\gamma_{R_i E_2} = \frac{a_2 \rho_r |h_{R_i E}|^2}{a_1 \rho_r |h_{R_i E}|^2 + d_{RE}^\alpha} \quad (15)$$

III. SOP ANALYSIS

In this part, the SOPs of the cooperative NOMA system using three kinds of relay selection schemes are studied.

To get the SOP for every user, channel statistics for the users and *Eve* are analyzed primarily. Combined with (5) and (12), the CDF of SINR from BS to D_1 can be written as

$$\begin{aligned} F_{\gamma_1}(x) &= \Pr(\min(\gamma_{R_i,D_1}, \gamma_{D_1}) < x) \\ &= 1 - \Pr(\gamma_{R_i,D_1} > x) \Pr(\gamma_{D_1} > x) \\ &= 1 - \Pr\left(\frac{a_1 \rho_s |h_{SR_i}|^2}{d_{SR}^\alpha} > x\right) \Pr\left(\frac{a_1 \rho_r |h_1|^2}{d_{RD_1}^\alpha} > x\right) \\ &= 1 - e^{-\frac{Ax}{a_1}}, \end{aligned} \quad (16)$$

where $\gamma_1 = \min\{\gamma_{R_i,D_1}, \gamma_{D_1}\}$, $A = \frac{d_{SR}^\alpha}{\rho_s g_{SR_i}} + \frac{d_{RD_1}^\alpha}{\rho_r g_1}$.

In similar, the CDF of SINR from BS to D_2 is given as

$$F_{\gamma_2}(x) = \begin{cases} 1 - e^{-\frac{Bx}{(a_2 - a_1)x}} & x \leq \frac{a_2}{a_1} \\ 1 & x > \frac{a_2}{a_1}, \end{cases} \quad (17)$$

where $\gamma_2 = \min\{\gamma_{R_i,D_2}, \gamma_{D_1,D_2}, \gamma_{D_2}\}$, and $B = \frac{d_{SR}^\alpha}{\rho_s g_{SR_i}} + \frac{d_{RD_1}^\alpha}{\rho_r g_1} + \frac{d_{RD_2}^\alpha}{\rho_r g_2}$.

For the signals received in *Eve* ($BS \rightarrow E, R_i \rightarrow E$), the SC algorithm is employed. Then, according to (5), (7) and (14), the CDF of γ_{E_1} is expressed as

$$\begin{aligned} F_{\gamma_{E_1}}(x) &= \Pr(\max(\gamma_{SE_1}, \min(\gamma_{R_i,D_1}, \gamma_{R_i E_1}) < x)) \\ &= \Pr(\gamma_{SE_1} < x) (1 - \Pr(\min(\gamma_{R_i,D_1}, \gamma_{R_i E_1}) > x)) \\ &= \Pr(\gamma_{SE_1} < x) (1 - \Pr(\gamma_{R_i,D_1} > x) \Pr(\gamma_{R_i E_1} > x)) \\ &= \left(1 - e^{-\frac{d_{SE}^\alpha x}{a_1 g_{SE}}}\right) \left(1 - e^{-\frac{x}{a_1} \left(\frac{d_{SR}^\alpha}{g_{SR_i}} + \frac{d_{RE}^\alpha}{g_{R_i E}}\right)}\right) \\ &= \left(1 - e^{-\frac{Ex}{a_1}}\right) \left(1 - e^{-\frac{Cx}{a_1}}\right), \end{aligned} \quad (18)$$

where $C = \frac{d_{SR}^\alpha}{\rho_s g_{SR_i}} + \frac{d_{RE}^\alpha}{\rho_r g_{R_i E}}$, and $E = \frac{d_{SE}^\alpha}{\rho_s g_{SE}}$.

The PDF of γ_{E_1} can be obtained as

$$f_{\gamma_{E_1}}(x) = \frac{E}{a_1} e^{-\frac{Ex}{a_1}} + \frac{C}{a_1} e^{-\frac{Cx}{a_1}} - \frac{D}{a_1} e^{-\frac{Dx}{a_1}}, \quad (19)$$

where $D = \frac{d_{SR}^\alpha}{\rho_s g_{SR_i}} + \frac{d_{RE}^\alpha}{\rho_r g_{R_i E}} + \frac{d_{SE}^\alpha}{\rho_s g_{SE}}$.

Referring to the derivation of γ_{E_1} , the PDF of γ_{E_2} can be derived as

$$\begin{aligned} f_{\gamma_{E_2}}(x) &= \frac{E a_2}{(a_2 - a_1 x)^2} e^{-\frac{Ex}{a_2 - a_1 x}} \\ &\quad + \frac{C a_2}{(a_2 - a_1 x)^2} e^{-\frac{Cx}{a_2 - a_1 x}} \\ &\quad - \frac{D a_2}{(a_2 - a_1 x)^2} e^{-\frac{Dx}{a_2 - a_1 x}}. \end{aligned} \quad (20)$$

A. SOP FOR RRS

The SOP is a very important benchmark to evaluate systematic secure performance, we can formulate it as [40]

$$P_{out} = \Pr\left(\lceil C_{D_j} - C_{E_j} \rceil^+ < R_{th}\right), \quad (21)$$

where $\lceil X \rceil^+ = \max\{X, 0\}$, R_{th} is the threshold of rate.

The SOP for RRS can be rewritten as

$$\begin{aligned} SOP_{RRS} &= \Pr\left(\lceil C_{D_1} - C_{E_1} \rceil^+ < R_{th_1} \text{ or} \right. \\ &\quad \left. \lceil C_{D_2} - C_{E_2} \rceil^+ < R_{th_2}\right) \\ &= 1 - \Pr\left(\lceil C_{D_1} - C_{E_1} \rceil^+ > R_{th_1}, \right. \\ &\quad \left. \lceil C_{D_2} - C_{E_2} \rceil^+ > R_{th_2}\right) \\ &= 1 - \Pr\left(\frac{1 + \gamma_1}{1 + \gamma_{E_1}} > \varepsilon_1, \frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2\right), \end{aligned} \quad (22)$$

where $\varepsilon_j = 2^{R_{th_j}}$ with R_{th_j} being the target rates of D_j .

Note that the variables γ_j, γ_{E_j} in (22) are related, acquiring an accurate expression of SOP is difficult. Therefore, the upper bound of SOP is given using the basic probability theory, (22) can be rewritten as

$$\begin{aligned}
 &SOP_{RRS} \\
 &\leq \min \left\{ 1, 2 - \Pr \left(\frac{1 + \gamma_1}{1 + \gamma_{E_1}} > \varepsilon_1 \right) \right. \\
 &\quad \left. - \Pr \left(\frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2 \right) \right\} \\
 &= \min \left\{ 1, \underbrace{\Pr \left(\frac{1 + \gamma_1}{1 + \gamma_{E_1}} < \varepsilon_1 \right)}_{p_1^{out}} + \underbrace{\Pr \left(\frac{1 + \gamma_2}{1 + \gamma_{E_2}} < \varepsilon_2 \right)}_{p_2^{out}} \right\}. \tag{23}
 \end{aligned}$$

The term p_j^{out} in (23) represents the SOP for RRS at D_j and can be calculated as

$$\begin{aligned}
 p_j^{out} &= \Pr (\gamma_j < \varepsilon_j (1 + \gamma_{E_j}) - 1) \\
 &= \int_0^\infty f_{\gamma_{E_j}}(x) F_{\gamma_j}(\varepsilon_j (1 + x) - 1) dx. \tag{24}
 \end{aligned}$$

Then, on the basis of (16), (19) and (24), p_1^{out} can be obtained as (25), shown at the bottom of the page.

Take full advantage of (24), the SOP for RRS at D_2 can be written as

$$\begin{aligned}
 p_2^{out} &= \int_0^\mu f_{\gamma_{E_2}}(x) F_{\gamma_2}(\varepsilon_2(1+x) - 1) dx \\
 &\quad + \int_\mu^\infty f_{\gamma_{E_2}}(x) dx, \tag{26}
 \end{aligned}$$

where $\mu = \frac{1}{a_1 \varepsilon_2} - 1$.

With the combination of (17), (20), (26) and the Gaussian-Chebyshev quadrature method, the SOP for RRS at D_2 is given by (27), which is shown at the bottom of the page, where $\phi_t = \cos \left(\frac{2t-1}{2N} \pi \right), t \in \{l, m, n\}$, and

$$\begin{aligned}
 \varphi_1(x) &= \frac{1}{(a_2 - a_1 x)^2} e^{-\frac{Ex}{(a_2 - a_1 x)}} e^{-\frac{B(\varepsilon_2 + \varepsilon_2 x - 1)}{(a_2 - a_1(\varepsilon_2(1+x) - 1))}}, \\
 \varphi_2(x) &= \frac{1}{(a_2 - a_1 x)^2} e^{-\frac{Cx}{(a_2 - a_1 x)}} e^{-\frac{B(\varepsilon_2 + \varepsilon_2 x - 1)}{(a_2 - a_1(\varepsilon_2(1+x) - 1))}}, \\
 \varphi_3(x) &= \frac{1}{(a_2 - a_1 x)^2} e^{-\frac{Dx}{(a_2 - a_1 x)}} e^{-\frac{B(\varepsilon_2 + \varepsilon_2 x - 1)}{(a_2 - a_1(\varepsilon_2(1+x) - 1))}}.
 \end{aligned}$$

Combining (23), (25) and (27), the SOP for RRS is shown by (28), shown at the bottom of the page.

B. SOP FOR SRS

In this part, we consider the SRS scheme for HD-based cooperative NOMA. BS can randomly select a relay as its auxiliary to transpond the messages. Maximizing the minimum data transmission rate D_2 is the main idea of SRS method. What's more, the range of data rate for D_2 is dominant by three different data rates: i) the transmission rate for the relay R_i to decode messages x_2 , ii) the transmission rate for D_1 to decode messages x_2 . iii) the transmission rate for D_2 to decode messages x_2 . In relaying networks, the SRS scheme activates a relay, which can be expressed as

$$\begin{aligned}
 i_{SRS}^* &= \arg \max_i \left\{ \min \left\{ \log(1 + \gamma_{R_i, D_2}), \right. \right. \\
 &\quad \left. \left. \log(1 + \gamma_{D_1, D_2}), \log(1 + \gamma_{D_2}) \right\}, i \in S_R^1 \right\}, \tag{29}
 \end{aligned}$$

where S_R^1 reveals the amount of relays in the network. Pay attention that the HD-based SRS scheme inherits the advantage of guaranteeing the data rate of D_2 ,

$$\begin{aligned}
 p_1^{out} &= \int_0^\infty f_{\gamma_{E_1}}(x) F_{\gamma_1}(\varepsilon_1(1+x) - 1) dx \\
 &= 1 - \int_0^\infty \left(\frac{E}{a_1} e^{-\frac{Ex + A(\varepsilon_1(1+x) - 1)}{a_1}} + \frac{C}{a_1} e^{-\frac{Cx + A(\varepsilon_1(1+x) - 1)}{a_1}} - \frac{D}{a_1} e^{-\frac{Dx + A(\varepsilon_1(1+x) - 1)}{a_1}} \right) dx \\
 &= 1 - \left(\frac{E}{E + A\varepsilon_1} + \frac{C}{C + A\varepsilon_1} - \frac{D}{D + A\varepsilon_1} \right) e^{-\frac{A(\varepsilon_1 - 1)}{a_1}}. \tag{25}
 \end{aligned}$$

$$\begin{aligned}
 p_2^{out} &= 1 - \int_0^\mu f_{\gamma_{E_2}}(x) e^{-\frac{B(\varepsilon_2 + \varepsilon_2 x - 1)}{(a_2 - a_1(\varepsilon_2(1+x) - 1))}} dx \approx 1 - \frac{E a_2 \mu \pi}{2N} \sum_{l=0}^N \sqrt{1 - \phi_l^2} \varphi_1 \left(\frac{\mu \phi_l + \mu}{2} \right) \\
 &\quad - \frac{C a_2 \mu \pi}{2N} \sum_{m=0}^N \sqrt{1 - \phi_m^2} \varphi_2 \left(\frac{\mu \phi_m + \mu}{2} \right) + \frac{D a_2 \mu \pi}{2N} \sum_{n=0}^N \sqrt{1 - \phi_n^2} \varphi_3 \left(\frac{\mu \phi_n + \mu}{2} \right). \tag{27}
 \end{aligned}$$

$$\begin{aligned}
 SOP_{RRS} &= \min \left\{ 1, 2 - \frac{E}{E + A\varepsilon_1} e^{-\frac{A(\varepsilon_1 - 1)}{a_1}} - \frac{C}{C + A\varepsilon_1} e^{-\frac{A(\varepsilon_1 - 1)}{a_1}} + \frac{D}{D + A\varepsilon_1} e^{-\frac{A(\varepsilon_1 - 1)}{a_1}} - \frac{a_2 \mu \pi}{2N} \times \right. \\
 &\quad \left. \left(E \sum_{l=0}^N \sqrt{1 - \phi_l^2} \varphi_1 \left(\frac{\mu \phi_l + \mu}{2} \right) + C \sum_{m=0}^N \sqrt{1 - \phi_m^2} \varphi_2 \left(\frac{\mu \phi_m + \mu}{2} \right) - D \sum_{n=0}^N \sqrt{1 - \phi_n^2} \varphi_3 \left(\frac{\mu \phi_n + \mu}{2} \right) \right) \right\}. \tag{28}
 \end{aligned}$$

where applications for lower target data rate can be implemented.

In accordance with the above investigations, Ξ_1 denotes that either the relay i_{TRS}^* or any of the legal users is unable to decode x_2 safely. So, the SOP based on SRS scheme with HD can be obtained as follows,

$$\begin{aligned} SOP_{SRS} &= \Pr(\Xi_1) = \Pr\left(\left|S_R^1\right| = 0\right) \\ &= \prod_{i=1}^K \left(1 - \Pr\left(\frac{1 + \min(\gamma_{R_i,D_2}, \gamma_{D_1,D_2}, \gamma_{D_2})}{1 + \gamma_{E_2}} > \varepsilon_2\right)\right) \\ &= \prod_{i=1}^K \left(1 - \Pr\left(\frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2\right)\right), \end{aligned} \quad (30)$$

where $|S_R^1|$ denotes the size of S_R^1 .

Substituting (27) into (30), the SOP for SRS scheme can be obtained, that is shown by (31) at the bottom of the page.

C. SOP FOR TRS

For HD-based cooperative NOMA, TRS consists of two main periods. In the first period, the objective data rate of D_2 is met. In the second period, we expect to make the data transmission rate of D_1 as high as possible under the condition that the data transmission rate of D_2 is satisfied. Therefore, the first period activates the relays that meet the following conditions,

$$\begin{aligned} S_R^2 = \{ \log(1 + \gamma_{R_i,D_2}) \geq R_{D_2}, \log(1 + \gamma_{D_1,D_2}) \geq R_{D_2}, \\ \log(1 + \gamma_{R_i,D_2}) \geq R_{D_2}, 1 \leq i \leq K \}, \end{aligned} \quad (32)$$

where S_R^2 denotes these relays satisfying the objective data rate of D_2 in the first stage.

For all relays from S_R^2 , the second period chooses a relay to transmit messages and maximizes the data rate of D_1 , the selected relay is

$$\begin{aligned} i_{TRS}^* = \arg \max_i \{ \min \{ \log(1 + \gamma_{R_i,D_1}), \\ \log(1 + \gamma_{D_1}) \}, i \in S_R^2 \}. \end{aligned} \quad (33)$$

$$\begin{aligned} SOP_{SRS} = \Pr(\Xi_1) \approx \left(1 - \frac{a_2 \mu \pi}{2N} \left(E \sum_{l=0}^N \sqrt{1 - \phi_l^2} \varphi_1 \left(\frac{\mu \phi_l + \mu}{2} \right) \right. \right. \\ \left. \left. - C \sum_{m=0}^N \sqrt{1 - \phi_m^2} \varphi_2 \left(\frac{\mu \phi_m + \mu}{2} \right) + D \sum_{n=0}^N \sqrt{1 - \phi_n^2} \varphi_3 \left(\frac{\mu \phi_n + \mu}{2} \right) \right) \right)^K. \end{aligned} \quad (31)$$

$$T_1 = \min \left\{ 1, \left[\frac{\left(\frac{E}{E+A\varepsilon_1} - \frac{C}{C+A\varepsilon_1} + \frac{D}{D+A\varepsilon_1} \right) e^{-\frac{A(\varepsilon_1-1)}{a_1}}}{\frac{a_2 \mu \pi}{2N} \left(E \sum_{l=0}^N \sqrt{1 - \phi_l^2} \varphi_1 \left(\frac{\mu \phi_l + \mu}{2} \right) + C \sum_{m=0}^N \sqrt{1 - \phi_m^2} \varphi_2 \left(\frac{\mu \phi_m + \mu}{2} \right) - D \sum_{n=0}^N \sqrt{1 - \phi_n^2} \varphi_3 \left(\frac{\mu \phi_n + \mu}{2} \right) \right)} \right]^i \right\}. \quad (37)$$

As can be seen from the above explanations, excepting for guaranteeing the data rate of D_2 , the TRS scheme based on HD can support D_1 to perform some background tasks.

It is worth noting that the total SOP events can be classified as

$$SOP_{TRS} = \Pr(\Xi_1) + \Pr(\Xi_2), \quad (34)$$

where Ξ_2 means that the relaying i_{TRS}^* , D_1 and D_2 can successfully decode x_2 , while the i_{TRS}^* and D_1 cannot successfully decode x_1 . Considering the analysis of the second period, $\Pr(\Xi_2)$ is expressed as

$$\Pr(\Xi_2) = \sum_{i=1}^K \underbrace{\Pr\left(\frac{1 + \gamma_1}{1 + \gamma_{E_1}} < \varepsilon_1 \mid |S_R^2| = i\right)}_{T_1} \underbrace{\Pr\left(|S_R^2| = i\right)}_{T_2}, \quad (35)$$

where $|S_R^2|$ represents the value of S_R^2 .

Because of the mathematical intractability in (22), T_1 can be given as

$$\begin{aligned} T_1 &= 1 - \Pr\left(\frac{1 + \gamma_1}{1 + \gamma_{E_1}} > \varepsilon_1 \mid |S_R^2| = i\right) \\ &= \left[1 - \frac{\Pr\left(\frac{1 + \gamma_1}{1 + \gamma_{E_1}} > \varepsilon_1, \frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2\right)}{\Pr\left(\frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2\right)} \right]^i. \end{aligned} \quad (36)$$

So, the term T_1 can be rewritten as (37) by substituting (25) and (27) into (36), it is shown at the bottom of the page.

Moreover, there exist i relays in S_R^2 , so the corresponding probability T_2 is calculated by

$$\begin{aligned} T_2 &= \binom{K}{i} \left(\Pr\left(\frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2\right) \right)^i \\ &\quad \times \left(1 - \Pr\left(\frac{1 + \gamma_2}{1 + \gamma_{E_2}} > \varepsilon_2\right) \right)^{K-i} \\ &= \binom{K}{i} (1 - p_2^{out})^i (p_2^{out})^{K-i}. \end{aligned} \quad (38)$$

Combining (31), (35), (37) and (38) and employing some arithmetical operations, the SOP for TRS scheme can be

expressed as

$$SOP_{TRS} = \sum_{i=0}^K \binom{K}{i} (1 - p_2^{out})^i (p_2^{out})^{K-i} \times \min \left(1, \left[\frac{p_1^{out}}{1 - p_2^{out}} \right]^i \right). \quad (39)$$

IV. ASYMPTOTIC SOP ANALYSIS

To gain deeper insights, the asymptotical SOPs of cooperative NOMA over Rayleigh fading channel are analyzed under these RS schemes. As $\rho \rightarrow \infty$ ($\rho_s = \rho_r$), specifically, the SOP of cooperative NOMA systems under each RS scheme depends on far user D_2 when $\gamma_2 \rightarrow \infty$. The asymptotical SOP for cooperative NOMA is shown as

$$ASOP_{RRS} \approx \Pr \left([C_{D_2} - C_{E_2}]^+ < R_{th_2} \right). \quad (40)$$

Substituting (27) into (40), the asymptotic SOP for RRS scheme when $\rho \rightarrow \infty$ can be obtained by

$$ASOP_{RRS} = 1 - \frac{a_2 \mu \pi}{N} \left\{ \sum_{l=0}^N \frac{2E \sqrt{1 - \phi_l^2}}{(2a_2 - a_1 \mu (\phi_l + 1))^2} + \sum_{m=0}^N \frac{2C \sqrt{1 - \phi_m^2}}{(2a_2 - a_1 \mu (\phi_m + 1))^2} - \sum_{n=0}^N \frac{2D \sqrt{1 - \phi_n^2}}{(2a_2 - a_1 \mu (\phi_n + 1))^2} \right\}. \quad (41)$$

From the asymptotic expression of SOP, it can be seen that there exists secure performance floor in cooperative NOMA system, which depends on the NOMA protocol. The main cause for this situation is that the realizable data rate of far user D_2 is restricted by the power distribution coefficient, a_2/a_1 . However, there is no such restriction to realize the data rate in *Eve*.

Based on (31), we observe that $\Pr(\Xi_1)$ tends to a constant. Therefore, the asymptotic SOP under SRS scheme is given by

$$ASOP_{SRS} = \left\{ 1 - \frac{a_2 \mu \pi}{N} \left\{ \sum_{l=0}^N \frac{2E \sqrt{1 - \phi_l^2}}{(2a_2 - a_1 \mu (\phi_l + 1))^2} + \sum_{m=0}^N \frac{2C \sqrt{1 - \phi_m^2}}{(2a_2 - a_1 \mu (\phi_m + 1))^2} - \sum_{n=0}^N \frac{2D \sqrt{1 - \phi_n^2}}{(2a_2 - a_1 \mu (\phi_n + 1))^2} \right\} \right\}^K. \quad (42)$$

Furthermore, taking into account the second stage, we have

$$\Pr(\Xi_2) = 0, \rho \rightarrow \infty. \quad (43)$$

According to the above analysis, the asymptotic SOP under TRS scheme is similar to SRS scheme. That is to say, $ASOP_{TRS} = ASOP_{SRS}$.

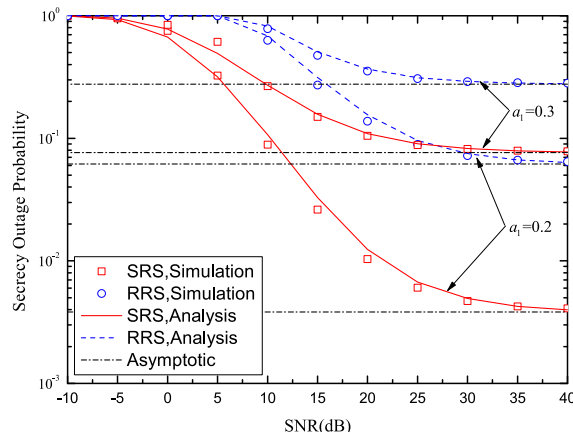


FIGURE 2. SOP versus the transmit SNR for RRS and SRS schemes with $K = 2, \alpha = 3, R_{th_1} = 2$ and $R_{th_2} = 0.5$.

By comparing asymptotic SOP under RRS scheme with SRS and TRS schemes, we find that the SRS and TRS schemes prominently improve the secrecy outage performance, and the interesting discovery is that increasing the amount of relays can further enhance the security performance.

V. NUMERICAL RESULTS

In this section, theoretical and practical simulation results are provided. The abbreviation for the bit-per-channel-use is BPCU. Combined with complexity and exactitude, we set up tradeoff parameter: $N = 30$.

Fig. 2 is drawn to describe the SOP of cooperative NOMA under RRS and SRS schemes for different power distribution coefficients with $K = 2, \alpha = 3, d_{SR} = 0.5, d_{RD_1} = 0.3, d_{RD_2} = 0.5, d_{SE} = 0.8, d_{RE} = 0.6, R_{th_1} = 2$ and $R_{th_2} = 0.5$, where $a_2 > a_1$ and $a_2 = 1 - a_1$. The blue circles and dash curves indicate the accurate SOP of RRS scheme for HD-based NOMA. The red squares and solid curves are the SRS strategy for cooperative NOMA, as can be seen from the accurate result obtained in (31). The curves of theoretical SOP coincide with the statistical simulation results. No matter what SNR situation is, the performance of SRS strategy is preferable to RRS strategy. Moreover, the SOP of the HD-based SRS and RRS schemes under $a_1 = 0.2$ and $a_2 = 0.8$ outperforms the SRS and RRS schemes under $a_1 = 0.3$ and $a_2 = 0.7$, respectively. Another phenomenon can be clearly obtained that HD-based NOMA RRS scheme under $a_1 = 0.2$ and $a_2 = 0.8$ is superior to SRS scheme under $a_1 = 0.3$ and $a_2 = 0.7$ in high SNR range. The reason for this situation is that the power distribution coefficient has a great influence in HD-based RS strategies. In addition, the simulation results also show that the security requirements of the nearby user D_1 have no effect on the security performance layer, which also proves the conclusion of the approximate SOP analyzed in the previous discussion.

Fig. 3 depicts the SOP of cooperative NOMA under RRS and TRS schemes for different power

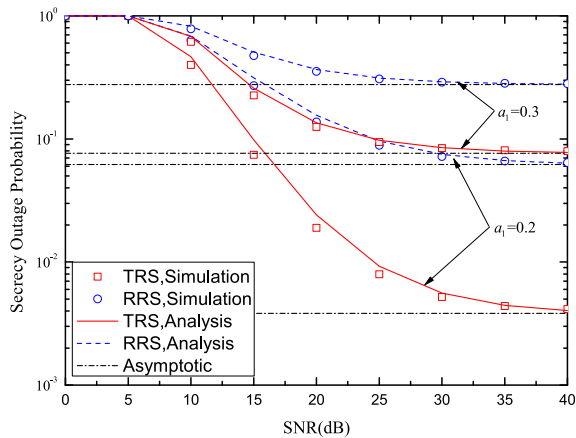


FIGURE 3. SOP versus the transmit SNR for RRS and TRS schemes with $K = 2$, $\alpha = 3$, $R_{th1} = 2$ and $R_{th2} = 0.5$.

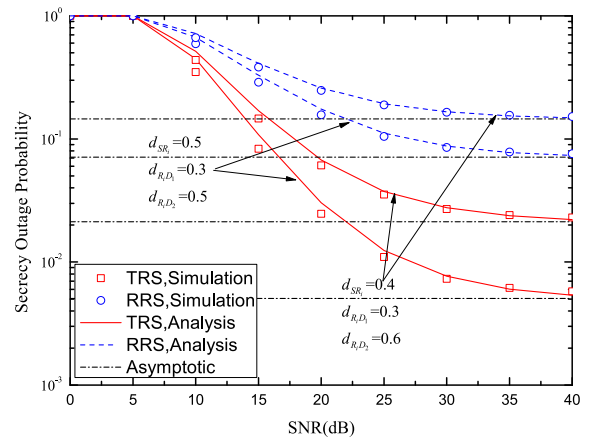


FIGURE 5. SOP versus the transmit SNR for RRS and TRS schemes for the different distances with $R_{th1} = 1$ and $R_{th2} = 0.3$.

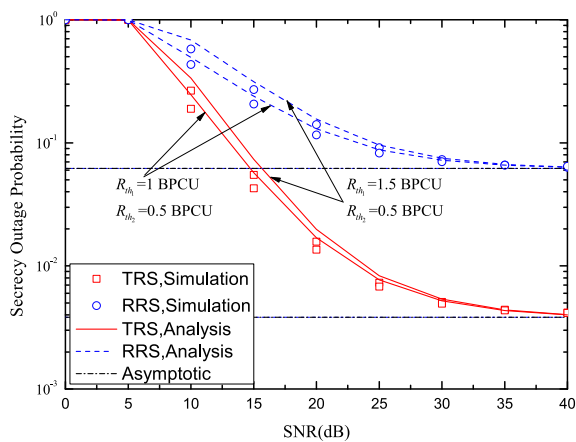


FIGURE 4. SOP versus the transmit SNR for RRS and TRS schemes for the different target rates with $a_1 = 0.2$ and $K = 2$.

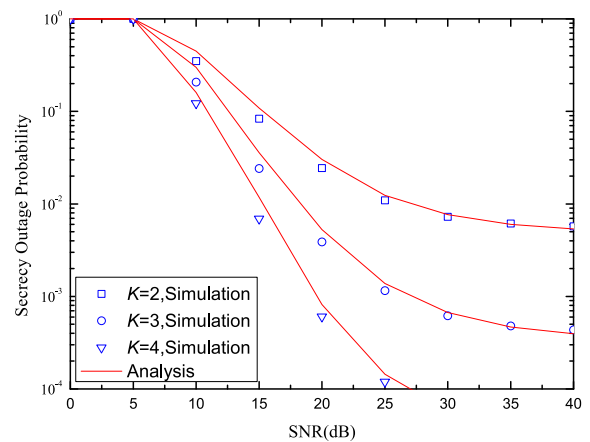


FIGURE 6. SOP versus the transmit SNR for TRS schemes with $K = 2, 3, 4$, $R_{th1} = 1$ and $R_{th2} = 0.3$.

distribution coefficients. The red lines represent TRS scheme for cooperative NOMA, and are consistent with the results obtained in (39). According to the analysis results, the TRS strategy can strengthen security performance. Similarly, the SOP of TRS and RRS schemes under $a_1 = 0.2$ and $a_2 = 0.8$ outperforms the TRS and RRS schemes under $a_1 = 0.3$ and $a_2 = 0.7$, respectively. Moreover, when $SNR < 25$ dB, the security performance of RRS scheme with $a_1 = 0.2$ is inferior to the TRS scheme with $a_1 = 0.3$. When $SNR > 25$ dB, the security performance of RRS scheme with $a_1 = 0.2$ begins to improve and surpasses the security performance of TRS scheme with $a_1 = 0.3$. Therefore, power distribution coefficient has a great influence on the security performance. It is also worth noting that the SOP is saturated in high SNR, and the target confidentiality rate of legal user D_2 can determine the lower performance. The primary cause for this phenomenon is that the NOMA protocol limits the available data rate for weak user D_2 .

In Fig. 4, we compare the SOP using RRS and TRS strategies with different target transmission rates. An interesting observation is that transforming the NOMA user's target rate can affect the security outage behaviors for HD-based RRS

and TRS schemes. As the target rate value reduces, the two kinds of schemes provide better outage performance, but the advantage fades away in high SNR range. Even if an effective RS scheme is implemented, there also exists secrecy performance floor. This is because the application of these two schemes does not eliminate the limitations (e.g., a_2/a_1) imposed by the NOMA protocol.

In Fig. 5, the SOP of cooperative NOMA under RRS and TRS schemes for different distances with $K = 2$, $a_1 = 0.2$, $a_2 = 0.8$, $\alpha = 3$, $d_{SE} = 0.6$, $d_{RE} = 0.4$, $R_{th1} = 1$ and $R_{th2} = 0.3$. This paper normalizes the distances for d_{SR} and d_{RD_2} , where $d_{RD_1} < d_{RD_2}$ and $d_{RD_1} + d_{RD_2} = 1$, because D_1 is the nearby user, whereas D_2 is the far user. It is observed that the security performance of TRS scheme is superior to RRS scheme when changing the distance. Compared with RRS scheme, there are more obvious variations for TRS scheme with different distances on security performance. Therefore, the distance from BS to R_i and from R_i to D_j has a significant impact on the secure outage performance for HD-based systems. Similarly, the SOP of cooperative NOMA can be influenced by d_{SE} and d_{RE} .

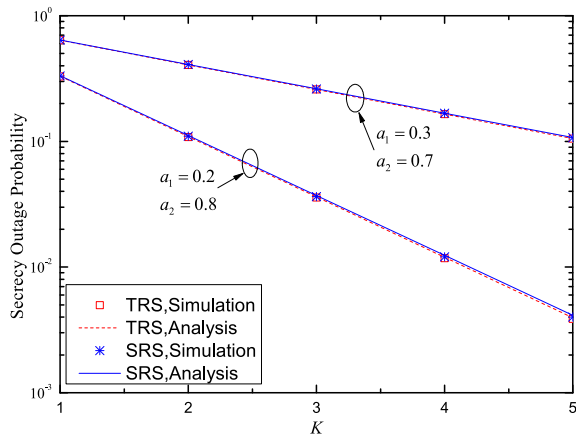


FIGURE 7. SOP versus K for SRS and TRS schemes with $R_{th_1} = 1$ and $R_{th_2} = 0.5$.

Fig. 6 paints the SOP employing TRS scheme when the number of relays is $K = 2, 3, 4$. The results show that the quantity of relays in this model has great effect on the performance of HD-based TRS schemes. When the number of relays increases, the RS schemes can achieve the lower outage probability. The reason is that the number of relays is positively correlated with diversity gain, thus it can improve the reliability of the cooperative networks.

Fig. 7 shows the SOP for both SRS and TRS schemes with respect to the number of relays K in high SNR region. It is observed that the analytic curves are precisely consistent with the simulated results. It can be concluded that the SOP using RS schemes reduces as the quantity of relays K increases. The security performance is improved due to the application of the efficient RS schemes that take advantage of the diversity of relaying networks. Moreover, from the analysis in section IV and expressions (31), (39), the SOP of both SRS and TRS schemes is coincident on account of $p_1^{out} = 0$ in strong SNR. Another conclusion is that the SOP of the RS schemes becomes smaller when the increasing of a_2/a_1 distinctly.

VI. CONCLUSION

This paper has studied the security performance for cooperative HD-based NOMA IoT systems over Rayleigh-distributed under the influence of different relay selection methods. The closed-form formulae of SOP for two users are derived. Further analysis shows that the SRS/TRS scheme can achieve the best secure performance, and RRS strategy may increase the SOP compared with SRS/TRS strategy. The security performance can be enhanced by augmenting the quantity of relays. Whereas, it is pointed out that due to the adoption of NOMA system, each RS scheme exists secrecy performance floor that cannot be deleted by RS schemes and power allocation strategy.

REFERENCES

[1] X. Chen, L. Guo, X. Li, C. Dong, J. Lin, and P. T. Mathiopoulos, "Secrecy rate optimization for cooperative cognitive radio networks aided by a wireless energy harvesting jammer," *IEEE Access*, vol. 6, pp. 34127–34134, 2018.

[2] Y. Cao, N. Zhao, G. Pan, Y. Chen, L. Fan, M. Jin, and M.-S. Alouini, "Secrecy analysis for cooperative NOMA networks with multi-antenna full-duplex relay," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5574–5587, Aug. 2019.

[3] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[4] F. Jameel, S. Wyne, G. Kaddoum, and T. Q. Duong, "A comprehensive survey on cooperative relaying and jamming strategies for physical layer security," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2734–2771, 3rd Quart., 2019.

[5] J.-H. Lee, "Full-duplex relay for enhancing physical layer security in multi-hop relaying systems," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 525–528, Apr. 2015.

[6] A. Mukherjee, "Physical-layer security in the Internet of Things: Sensing and communication confidentiality under resource constraints," *Proc. IEEE*, vol. 103, no. 10, pp. 1747–1761, Oct. 2015.

[7] L. Qing, H. Guanyao, and F. Xiaomei, "Physical layer security in multi-hop AF relay network based on compressed sensing," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1882–1885, Sep. 2018.

[8] A. Pandey, S. Yadav, T. Do, and R. Kharel, "Secrecy performance of cooperative cognitive AF relaying networks with direct links over mixed Rayleigh and double-Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, early access, Oct. 29, 2020, doi: 10.1109/TVT.2020.3034729.

[9] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-level performance evaluation of downlink non-orthogonal multiple access (NOMA)," in *Proc. IEEE 24th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 611–615.

[10] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[11] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[12] X. Li, Q. Wang, M. Liu, J. Li, H. Peng, J. Piran, and L. Li, "Cooperative wireless-powered NOMA relaying for B5G IoT networks with hardware impairments and channel estimation errors," *IEEE Internet Things J.*, early access, Oct. 9, 2020, doi: 10.1109/JIOT.2020.3029754.

[13] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[14] T. Nakamura, A. Benjebbour, Y. Kishiyama, S. Suyama, and T. Imai, "5G radio access: Requirements, concept and experimental trials," *IEICE Trans. Commun.*, vol. E98.B, no. 8, pp. 1397–1406, 2015.

[15] D. Do, T. Nguyen, K. M. Rabie, X. Li, and B. M. Lee, "Throughput Analysis of Multipair Two-Way Relaying Networks With NOMA and Imperfect CSI," *IEEE Access*, vol. 8, pp. 128942–128953, 2020.

[16] Y. Liu, Z. Qin, M. El-kashlan, Y. Gao, and L. Hanzo, "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1656–1672, Mar. 2017.

[17] H. Lei, J. Zhang, K.-H. Park, P. Xu, Z. Zhang, G. Pan, and M.-S. Alouini, "Secrecy outage of Max-Min TAS scheme in MIMO-NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 6981–6990, Aug. 2018.

[18] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under IQ imbalance," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12286–12290, Oct. 2020.

[19] K. Jiang, T. Jing, Y. Huo, F. Zhang, and Z. Li, "SIC-based secrecy performance in uplink noma multi-eavesdropper wiretap channels," *IEEE Access*, vol. 6, pp. 19664–19680, 2018.

[20] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[21] J. Choi, "Non-orthogonal multiple access in downlink coordinated two-point systems," *IEEE Commun. Lett.*, vol. 18, no. 2, pp. 313–316, Feb. 2014.

[22] J.-B. Kim and I.-H. Lee, "Non-orthogonal multiple access in coordinated direct and relay transmission," *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 2037–2040, Nov. 2015.

[23] Q. Li, P. Ren, and D. Xu, "Security enhancement and QoS provisioning for NOMA-based cooperative D2D networks," *IEEE Access*, vol. 7, pp. 129387–129401, 2019.

- [24] J. Men, J. Ge, and C. Zhang, "Performance analysis for downlink relaying aided non-orthogonal multiple access networks with imperfect CSI over Nakagami- m fading," *IEEE Access*, vol. 5, pp. 998–1004, 2017.
- [25] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.
- [26] J. Chen, L. Yang, and M.-S. Alouini, "Physical layer security for cooperative NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4645–4649, May 2018.
- [27] N. Dahi and N. Hamdi, "Relaying in non-orthogonal multiple access systems with simultaneous wireless information and power transfer," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2018, pp. 164–168.
- [28] Y. Zhang, H. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 930–933, May 2016.
- [29] B. Zheng, M. Wen, C.-X. Wang, X. Wang, F. Chen, J. Tang, and F. Ji, "Secure NOMA based two-way relay networks using artificial noise and full duplex," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1426–1440, Jul. 2018.
- [30] Y. Zou, X. Wang, and W. Shen, "Optimal relay selection for physical-layer security in cooperative wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2099–2111, Oct. 2013.
- [31] H. Lei, H. Zhang, I. S. Ansari, Z. Ren, G. Pan, K. A. Qaraqe, and M.-S. Alouini, "On secrecy outage of relay selection in underlay cognitive radio networks over Nakagami- m fading channels," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 614–627, Dec. 2017.
- [32] J. Zhao, Z. Ding, P. Fan, Z. Yang, and G. K. Karagiannidis, "Dual relay selection for cooperative NOMA with distributed space time coding," *IEEE Access*, vol. 6, p. 20440–20450, 2018.
- [33] Z. Ding, H. Dai, and H. Vincent Poor, "Relay selection for cooperative NOMA," *IEEE Wireless Commun. Lett.*, vol. 5, no. 4, pp. 416–419, Aug. 2016.
- [34] X. Li, H. Mengyan, Y. Liu, V. G. Menon, A. Paul, and Z. Ding, "IQ imbalance aware nonlinear wireless-powered relaying of B5G networks: Security and reliability analysis," *IEEE Trans. Netw. Sci. Eng.*, early access, Sep. 3, 2020, doi: 10.1109/TNSE.2020.3020950.
- [35] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Spatially random relay selection for Full/Half-duplex cooperative NOMA networks," *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3294–3308, Aug. 2018.
- [36] H. Zhang, H. Lei, I. S. Ansari, G. Pan, and K. A. Qaraqe, "Security performance analysis of DF cooperative relay networks over Nakagami- m fading channels," *KSII Trans. Internet Inf. Syst.*, vol. 11, no. 5, pp. 2416–2432, May 2017.
- [37] Y. Liu, Z. Qin, M. ElKashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2667–2680, Dec. 2017.
- [38] J. G. Proakis, *Digital Communications*, 4th ed. Boston, MA, USA: McGraw-Hill, 2001.
- [39] G. Liu, Z. Wang, J. Hu, Z. Ding, and P. Fan, "Cooperative NOMA Broadcasting/Multicasting for low-latency and high-reliability 5G cellular V2X communications," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7828–7838, Oct. 2019.
- [40] B. Li, Y. Zou, J. Zhou, F. Wang, W. Cao, and Y.-D. Yao, "Secrecy outage probability analysis of friendly jammer selection aided multiuser scheduling for wireless networks," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3482–3495, May 2019.



HUI LI received the B.Sc. degree in communication engineering from the School of Information Engineering, in 1999, and the M.Sc. degree in communication and information system and the Ph.D. degree in information and communication engineering from the Nanjing University of Science and Technology, in 2004 and 2008, respectively. He was a Visiting Scholar with Charles Darwin University, Australia, in 2013, and North Carolina A & T State University, in 2014. He is currently a Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo China. His research interests include wireless communications and intelligent signal processing.



YAPING CHEN received the B.Sc. degree in electronic information science and technology from the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo China, in 2019, where she is currently pursuing the M.Sc. degree in communication and information systems. Her research interests include physical layer security (PLS) and cooperative communication.



MINGFU ZHU received the B.Sc. degree from Tianjin University, in 2000, the M.Sc. degree from East China Normal University, in 2004, and the Ph.D. degree from the University of California, Los Angeles (UCLA), in 2007.

From 2011 to 2013, he was working as an Executive Director with Maynard Photoelectric Technology Company Ltd., Ningbo, China. Since 2013, he has founded and served as the Chairman of Hebei National Optoelectronics Technology Company Ltd., Hebei, China. He is currently the Chairman of Henan Chuangzhi Technology Company Ltd., and a General Manager of Henan Chuitian Technology Company Ltd., Hebei. His research interests include chip packaging and intelligent light development and manufacturing. With innovative ideas, intelligent lights are used to build IOL, integrate IOL into IoT, and upgrade to 5G IoT. By building scientific research platforms and manufacturing bases, 5G industry ecosystem is built to interact with upstream and downstream enterprises. He received several awards and achievements, which include the excellent builder for the socialist cause with Chinese characteristics, special government allowance under the State Council and leading talents in Science and Technology Innovation of National Ten Thousand Talents Plan and so on. He has served as a member of Henan CPPCC and a Vice President of the Henan Euro-American Alumni Association. He is also a President of the Henan Alumni Association of Tianjin University and the Director of the Henan Mechanical Engineering Society.



JIANGFENG SUN (Member, IEEE) received the M.S. degree in communication and information system from Zhengzhou University, in 2009. He is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications. He is currently a Lecturer with the School of College of Computer Science and Technology, Henan Polytechnic University. He has several papers published in journal and conferences. His current research interests include physical layer

security, cooperative communications, and performance analysis of fading channels.



DINH-THUAN DO received the B.S., M.Eng., and Ph.D. degrees in communications engineering from Vietnam National University (VNU-HCMC), in 2003, 2007, and 2013, respectively. From 2003 to 2009, he was a Senior Engineer with the VinaPhone Mobile Network. From 2009 to 2010, he was a Visiting Ph.D. Student with the Communications Engineering Institute, National Tsinghua University, Taiwan. His name and his achievements will be reported in special

book entitled *Young talents in Vietnam 2015-2020*. His research interests include signal processing in wireless communications networks, NOMA, full-duplex transmission, and energy harvesting. His publications include over 80 SCIE/SCI-indexed journal articles, over 45 SCOPUS-indexed journal articles, and over 50 international conference papers. He is sole author in one textbook and one book chapter. He was a recipient of the Golden Globe Award from Vietnam Ministry of Science and Technology, in 2015, (Top 10 most excellent scientist nationwide). He is currently serving as an Editor of *Computer Communications* (Elsevier), an Associate Editor of the *EURASIP Journal on Wireless Communications and Networking* (Springer), and an Editor of *KSII Transactions on Internet and Information Systems*.



VARUN G. MENON (Senior Member, IEEE) is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include the Internet of Things, fog computing and networking, underwater acoustic sensor networks, cyber psychology, hijacked journals, ad-hoc networks, and wireless sensor networks. He is a Distinguished Speaker of ACM Distinguished Speaker. He is also a Guest Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE SENSORS JOURNAL, the *IEEE Internet of Things Magazine*, and the *Journal of Supercomputing*. He is an Associate Editor of *IET Quantum Communications*. He is also an Editorial Board Member of the IEEE Future Directions: Technology Policy and Ethics.



SHYNU P. G. (Member, IEEE) received the Ph.D. degree in computer science from the Vellore Institute of Technology (VIT), Vellore, India, and the master's degree in engineering in computer science and engineering from the College of Engineering, Anna University, Chennai, India. He is currently working as an Associate Professor with the School of Information Technology and Engineering, VIT. He has published over 30 research papers in refereed international conferences and journals. His research interests include deep learning, cloud security and privacy, ad-hoc networks, and big data.

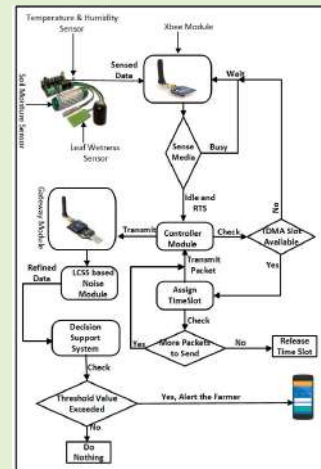
•••

Smart Sensing-Enabled Decision Support System for Water Scheduling in Orange Orchard

Rahim Khan¹, Muhammad Zakarya¹, Member, IEEE, Venki Balasubramanian², Member, IEEE, Mian Ahmad Jan¹, Senior Member, IEEE, and Varun G. Menon³, Senior Member, IEEE

Abstract—The scarcity of water resources throughout the world demands its optimum utilization in various sectors. Smart Sensing-enabled irrigation management systems are the ideal solutions to ensure the optimum utilization of water resources in the agriculture sector. This paper presents a wireless sensor network-enabled Decision Support System (DSS) for developing a need-based irrigation schedule for the orange orchard. For efficient monitoring of various in-field parameters, our proposed approach uses the latest smart sensing technology such as soil moisture, leaf-wetness, temperature and humidity. The proposed smart sensing-enabled test-bed was deployed in the orange orchard of our institute for approximately one year and successfully adjusted its irrigation schedule according to the needs and demands of the plants. Moreover, a modified Longest Common SubSequence (LCSS) mechanism is integrated with the proposed DSS for distinguishing multi-valued noise from the abrupt changing scenarios. To resolve the concurrent communication problem of two or more wasp-mote sensor boards with a common receiver, an enhanced RTS/CTS handshake mechanism is presented. Our proposed DSS compares the most recently refined data with pre-defined threshold values for efficient water management in the orchard. Irrigation activity is scheduled if water deficit criterion is met and the farmer is informed accordingly. Both the experimental and simulation results show that the proposed scheme performs better in comparison to the existing schemes.

Index Terms—Wireless sensor network, precision agriculture, irrigation management systems, DSS, RTS/CTS, LCSS.



I. INTRODUCTION

WORLDWIDE, water is a scarce resource and it requires considerable attention from the research community and industry to ensure its maximum utilization. Agriculture sector is one the water's main consumer because it consumes approximately 70% of the available water to fulfill the food requirements of the world fast growing population [1]. Generally, irrigation schedules are based on farmers experience, crop requirements, environmental conditions, and

soil properties. However, these traditional irrigation procedures are not efficient from the resource utilization perspective as a considerable amount of water is wasted. Due to the recent technological advancements, particularly sensors and actuators, it is possible to develop a smart sensing-enabled automated Decision Support System (DSS) that has the ability to identify water-deficit locations and irrigate those areas on priority basis if needed [2].

Manuscript received June 24, 2020; revised July 19, 2020; accepted July 24, 2020. Date of publication July 28, 2020; date of current version August 13, 2021. The associate editor coordinating the review of this article and approving it for publication was Dr. Hari P. Gupta. (Corresponding author: Mian Ahmad Jan.)

Rahim Khan, Muhammad Zakarya, and Mian Ahmad Jan are with the Department of Computer Science, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan (e-mail: rahimkhan@awkum.edu.pk; mohd.zakarya@awkum.edu.pk; mianjan@awkum.edu.pk).

Venki Balasubramanian is with the School of Engineering, Information Technology and Physical Sciences, Federation University Australia at Mount Helen Campus, Ballarat, VIC 3350, Australia.

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India.

Digital Object Identifier 10.1109/JSEN.2020.3012511

A network of smart sensing devices has the potentials to collect the real-time data for developing an automated Irrigation Management System (IMS) or DSS, that is also known as precision agriculture [3]. The DSS aims to provide the right resources at the right time, which has a direct correlation with the yield improvement of various crops. To realize this objective, smart sensing devices are deployed in agricultural fields where specialized sensors probe their surrounding phenomena such as soil moisture, soil temperature, pH, humidity, and leaf wetness, etc. These gathered phenomena are thoroughly observed by the DSS on a centralized device. Various agriculture-related activities are subject to these observations. There exist numerous studies in this context. In [4], a WSN-based remote water management system for agriculture

sector was implemented in Thailand. The proposed approach was deployed for five months in agricultural fields and various data mining techniques were used to analyze the captured data. However, the proposed approach completely neglected outliers, i.e., noisy data. A DSS-enabled irrigation prediction system was presented in [5] for optimizing the water scheduling of orange orchard. This system has the ability to predict the soil moisture of a particular region within the orchard by applying a hybrid of Support Vector Regression (SVR) and K-mean clustering algorithm on the gathered data. The proposed DSS emphasized on the refinement of captured data before it is being processed by the DSS. However, this system fails to distinguish multivariate noise from abrupt changing scenarios [6]. In [7], a WSN-enabled water management system was deployed in adjacent agricultural fields where different crops were grown. The proposed test-bed utilized both the historical data and variations in the climate values to devise an effective irrigation schedule. An Internet of Things (IoT) and neural network-based DSS was developed in [8] to predict the water-deficit locations. The proposed DSS is capable to precisely detect the required amount of water for irrigation. However, this approach does not consider the outliers in the gathered data.

In precision agriculture, outliers or noise is defined as unwanted data that severely affect the performance of an operational DSS. Usually, this problem occurs due to malfunctioning sensor nodes, interference, collision of packets, circuit failure, extreme pressure, high temperature, and other environmental conditions. As a result, refinement of sensed data prior to its processing by the concern DSS is a challenging issue. Moreover, in the case of shared media, collision of Request To Send / Clear To Send (RTS/CTS) packets and data packets is another challenging issue. Therefore, the development of a precise and accurate technology-assisted DSS is desperately needed to ensure maximum utilization of water in agriculture sector.

In this paper, a smart sensing-enabled DSS for the orange orchard is presented to resolve the aforementioned issues. In our proposed approach, the sensed data by the various smart sensing devices/nodes is processed using a refinement module to ensure accuracy and integrity of data at the destination. Moreover, every node is bounded to transmit its data only if the medium of communication is free. The main contributions of this paper are:

- 1) A smart sensing-enabled DSS is presented for proper management of the irrigation activities in agriculture sector. The proposed agricultural DSS is a need-based system that provides water to a particular area only if it is identified as water-deficit.
- 2) A modified LCSS mechanism is proposed that enables the proposed DSS to differentiate multivariate noise from the abrupt changing scenario.
- 3) An enhanced RTS/CTS mechanism is proposed to resolve the collision issue associated with concurrent communications. Before any transmission activity, every sensor node is bounded to use the classifier-based mechanism that ensures a collision-free communication between two or more operational devices.

The rest of the paper is organized as follows. In Section II, an overview of the literature is presented. In Section III, a detailed description of the proposed smart sensing-enabled DSS module and its deployment for orange orchard are presented. In Section IV, both experimental and simulation results are discussed in detail. Finally, concluding remarks are provided in Section V.

II. LITERATURE REVIEW

Precision agriculture is the technology-assisted farming, which is based on sensor-enabled monitoring, measurement and response generation via the DSS. The responses are generated based on the varying conditions of crops [9]. It enables the farmer to provide the right resources at the right time and right place to any crop [10]. In agriculture, water is an essential resource that is needed to bring forth the maximum potential of the agricultural fields. Moreover, it enables crops to make full use of other yield enhancing production factors [11]. In this section, a brief overview of various test-beds which are related to the proposed work is presented.

Keswani *et al.* [8] have presented an optimal Internet of things (IoTs) and neural network-based irrigation management system that has a one-hour prior prediction capability of water-deficit location(s). For this purpose, various sensors were deployed such as soil moisture, temperature, CO₂, light intensity, and humidity sensors. A hybrid DSS was proposed by Viani *et al.* [12] which was based on the fuzzy logic and farmer's experiences. Likewise, WSNs and General Packet Radio Services (GPRS) were used to form an optimal irrigation management system. Soil moisture sensors with controller modules were deployed in agricultural field(s) [13]. A machine learning and agronomist's encysted knowledge-based irrigation prediction system was proposed that concluded that Gradient Boosted Regression Trees (GBRT) was the best regression model with approximately 93% irrigation prediction accuracy [14]. Dursun and Ozden [15] presented an automatic drip irrigation management system for the cherry trees. A low-cost IoT system for smart irrigation was proposed by NK. Nawandar and Satpute [16]. The system capabilities include the estimation of irrigation schedule, neural-based decisions and remote monitoring. Similarly, a WSN-based irrigation management platform was presented that has the capacity to calculate the quantity of water needed for irrigating a specific area [17]. A novel watering management system, which is based on low-cost IoT components was presented by Khoa *et al.* [18]. Additionally, LoRa LPWAN technology was used for the transmission to ensure the best performance of the proposed system. In FLOW-AID project, WSNs were used to identify water-deficit locations, a situation where plants need water desperately [19]. In 2011, the Information and Communication Technology unit of Commonwealth Scientific and Industrial Research Organization (CSIRO) used various sensor nodes to recover the ecological integrity of Queensland's National Park [20]. Pardossi *et al.* [21] described a methodology of integrating Root Zone sensors with WSNs which is used to identify water deficit locations in agricultural fields. Harun *et al.* [22] described WSNs as an efficient tool to resolve both the decision-making and resource optimization

issues associated with technology-assisted farming. A need-based irrigation practice was presented by Abrishambaf *et al.* [23]. This system has the capacity to schedule the irrigation activity for lowest cost period by using various parameters to temperature, soil moisture, wind, precipitation forecast, and soil calculation. An IoT-based irrigation system was developed to automate the irrigation activity of crops using soil and environmental data [24]. Dong *et al.* [25] presented a pivot-based irrigation procedure to optimize the irrigation activity via wireless underground sensor networks.

III. PROPOSED LCSS-BASED DATA REFINEMENT MECHANISM

Data fusion or refinement of the WSNs capture data has become a dominant research area; as the majority of our daily-life activities are either partially or completely dependent on these DSS-based networks. In this section, a space free LCSS-based data fusion scheme is presented to enhance accuracy and precision level of the proposed agricultural DSS. The captured data of every device C_i , that is wasp-mote agricultural board in the proposed test-bed, is passed through the LCSS-based noise detection module that ensures the accuracy of the refined data. Moreover, the proposed fusion scheme has the capacity to distinguish outliers or noisy data from the abrupt changing scenarios, i.e., an abrupt change occurs if water directly interacts with soil moisture or leaf wetness sensors.

A. Sequence Matching: Definitions and Preliminaries

A sequence SQ_i is defined as a collection of related values, $a_1, a_2, \dots, a_n \in SQ_n$ where n represents length of the data set. The longest common subsequence (LCSS) is represented by $LCSS(k)$, where k defines length of the LCSS. Concatenation process in LCSS is defined as appending any two symbols X and Y to form a subsequence XY such that $X \& Y \in SQ_a$.

Definition 1: Any two values $X \in SQ_a$ and $Y \in SQ_b$ are considered as equal **iff** $\text{distance}(X, Y) \leq 0.05$ or $X \approx Y$.

definitions $LCSS(0, 0)$ is used to describe an empty LCSS.

Definition 2: A value $a_1 \in SQ_a$ is considered as a predecessor of another value $a_2 \in SQ_a$ in $LCSS$ **iff** $\text{index}(a_2 > a_1)$ and for both values \exists (a value $b_m \in SQ_b$ such that $a_1 \approx b_m$ and $a_2 \approx b_m$).

Definition 3: A value $a_1 \in SQ_a$ is not considered as a predecessor of another value $a_2 \in SQ_a$, that is $a_1 \notin LCSS_{\text{matched}}$, **iff** $\text{index}(a_2 < a_1)$ that is $1 > 2$. Although, for both values $a_1 \& a_2 \in SQ_a \exists$ ($b_y \& b_z \in SQ_b$ such that $a_1 \approx b_y \& a_2 \approx b_z \forall$ where y and z represent indexes information).

Definition 4: Similarity index of any two sequences or data sets $SQ_a \& SQ_b$ are higher **iff** length of their $LCSS_{\text{matched}} > \text{length}(SQ_{3 \times n/2} \& SQ_{3 \times m/2})$

Definition 5: The $LCSS(k)$ represents the LCSS of SQ_a and SQ_b **iff** $\forall (a_n \in SQ_a)$ there exist a $b_m \in SQ_b$ such that $a_n \approx b_m$ and $\exists (n' < n \text{ and } m' < m \text{ such that } LCSS(k - n', l - m')$ is generated by n' and m').

B. Computation of the LCSS

The proposed approach uses two different sequences of the same size n that is 10 in this case i.e., SQ_a for storing

current data values and SQ_b for previously transmitted data where $a = 1, 2, 3 \dots n$ and $b = 1, 2, 3 \dots m$. Every device $C_i \in WSN$ is bounded to store the collect data in SQ_a until $a = n$. Initial values for SQ_b is defined manually, once at the deployment stage of WSNs, and are updated according to collected data of sensor(s). For example, soil moisture sensor values are set according to the average values of three different soil moisture sensors which were deployed in dried soil i.e., 250Hz to 260Hz. Once, the network becomes fully operational i.e., sensors begin to probe the phenomena after the defined interval of time, that is 30 second in the deployed WSN infrastructure. In the proposed test-bed, every wasp-mote board C_i is bounded to store their captured data temporarily in sequence SQ_a until value of $a = 10$ and then send it to the gateway.

To refine this data, the proposed test-bed uses a modified form of LCSS and gap-free LCSS. LCSS is used to find the similarity indexes of the currently received data SQ_a and existing data SQ_b . Initially, a matching window control parameter δ is defined that is used to limit the matching window of a value in sequence SQ_a , i.e., $a = 1$, with another sequence SQ_b . In the proposed test-bed, the value of δ is set to three (3) which means that the first element of SQ_a is matched with at-most three elements of SQ_b **iff** these elements are not matched. In phase-I, the first element of SQ_a , i.e., $SQ_1 \in SQ_a$, is matched with element(s) of SQ_b , i.e., $b_1, b_2, \dots, b_\delta \in SQ_b$, such that either a match is found or maximum limit δ is reached. If first element $a_1 \in SQ_a$ matches with any element $b_{1 \text{ to } \delta} \in SQ_b$ then b_m is stored in class $LCSS_{\text{matched}}$ with its position information. However, if a_1 does not match with any element of SQ_b within the defined window δ then a_1 is ignored and subsequent element $a_2 \in SQ_a$ is processed. Likewise, second value $a_2 \in SQ_a$ is matched using similar approach with a slight modification that is its matching criteria with the SQ_b is subjected to the following conditions.

- 1) $a_2 \in SQ_a$ is matched with the first value of SQ_b **iff** $a_1 \in SQ_a$ does not have a matching value in the defined window, i.e., δ .
- 2) $a_2 \in SQ_a$ is matched with value of SQ_b that is stored after previously matched value i.e., $b_1 \in SQ_b$ **iff** $a_1 \in SQ_a \approx b_1 \in SQ_b$.

If $a_2 \in SQ_a$ has a match in SQ_b then matching value $b_{2 \text{ to } \delta} \in SQ_b$ is stored in class $LCSS_{\text{matched}}$ with its position information. For the remaining values of SQ_a , this process is repeatedly applied to compute their LCSS.

In phase-II, first value of SQ_a , i.e., $a_1 \in SQ_a$, is ignored **iff** $a_1 \in LCSS_{\text{matched}}$ and the required LCSS is not computed yet. The remaining values, i.e., $a_2, a_3, \dots, a_n \in SQ_a$, is considered as a refined data set which has nine values. Then, the aforementioned process, i.e., finding $LCSS_{\text{matched}}$ of SQ_a and SQ_b , is repeated. Both LCSSs, i.e., current $LCSS_{\text{matched}}$ and previous $LCSS_{\text{matched}}$, are compared and $LCSS$ with the maximum length is selected whereas other is discarded. This process is repeated until the required LCSS.

To understand this idea, consider two sequences SQ_a and SQ_b which contain data generated by the temperature sensor(s) i.e., $SS_n = 30 \ 34 \ 31 \ 30 \ 33 \ 35 \ 34 \ 30 \ 34 \ 32$ and SQ_m

= 33 30 32 34 30 33 34 30 34 32 where $n=m=10$ and $\delta = 3$. First value $30 \in SQ_a$ is matched with every value of SQ_b within the defined window $\delta = 3$; starting with the first, i.e., $32 \in SQ_b$. A match is encountered at the 2^{nd} position in SQ_b i.e., $30 = 30$. Value 30 is stored in $LCSS_{matched}$ with its position information. Second value $34 \in SQ_a$ is then matched with every value of SQ_b starting from the position 3^{rd} i.e., 31 in this case. However, 31 does not have a matching value in SQ_b within δ . Therefore, it is neglected and the subsequent value $31 \in SQ_b$ is processed which is matched with 3^{rd} value in SQ_b . 31 is stored with its location information in $LCSS_{matched}$. For the remaining values of SQ_a , this process is repeatedly applied until their LCSS is found. It is to be noted that phase-II is applicable only if the computed LCSS length is less than the length of $SQ_b/2$.

Lemma 1: $LCSS(p)$ represents the longest common subsequence of SQ_a & SQ_b of length n and m respectively **iff** $k \geq 1$ and $\exists (a_1 \dots p$ such that $a_1 \dots p \cong b_{1 \dots p} \therefore a_{1 \dots p} \in SQ_a$ and $b_{1 \dots p} \in SQ_b$) where $p \leq n$ & m .

Proof: Applying mathematical induction i.e., for $k = 1$ The $length(LCSS(1))$ is equal to 1 **iff** $a_1 \cong b_1$ where $a_1 \in SQ_a$ and $b_1 \in SQ_b$. (According to **Definition-4**). Hence, the lemma is true for $k = 1$.

Suppose that the lemma is true for $k - 1$ values. We need to prove that the lemma is true for k values. If $LCSS(n, m)$ represents the LCSS of SQ_a & SQ_b then $\exists (n' < n$ and $m' < m$ such that $LCSS(n', m')$ is the LCSS of SQ'_a & SQ'_b for $k - 1$ value).

According to our assumption,

$LCSS(n', m') = p'_1, p'_2, p'_3, \dots, p'_k$ such that $p' < p$ and $p'_1, p'_2, p'_3, \dots, p'_k \in SQ_a$ & $SQ_b \therefore a_n = b_m \therefore n' < n$ and $m' < m$.

Therefore, LCSS of SQ_a and SQ_b is of length k . $\therefore length(LCSS(n', m')) + length(LCSS(n, m)) = k$. Hence, it proves that k is the length of required ($LCSS(p)$).

Conversely, if $length(LCSS(n, m)) \geq k$ and $a_n = b_n$, where $a_n \in SQ_a$ and $b_m \in SQ_b$ then $\exists (n' < n$ and $m' < m$ such that $a_{n'} \cong b_{m'}$. Moreover, $length(LCSS(n', m')) = length(LCSS(n, m)) - 1 \geq k - 1$. Therefore, $LCSS(n', m')$ is the LCSS of $k - 1$ length data sets (By inductive Hypothesis). Hence, the proof i.e., $LCSS(p)$ represents the longest common subsequence of SQ_a & SQ_b . \square

C. Proposed Methodology: Classifier-Based Based RTS/CTS Handshake

To resolve one of the aforementioned issue, i.e., collision of RTS/CTS packets, a classifier-based scheduled RTS/CTS mechanism is presented. Every device $C_i \in IoTs$ shares its communication schedule T_s with the neighboring devices via a smaller scheduled-frame preferably after the deployment phase. The proposed communication scheme consists of two phases i.e., hop-count and classifier-based optimal neighbors discovery phases.

1) Hop-Count Discovery Phase: The base station module S_j broadcasts a scheduled-frame which contains a transmission schedule (T_s), hop-count (H_c) and back-off timer T_b . Moreover, the hop-count value is set to zero as base station is the ultimate destination for every device $C_i \in IoTs$ and

T_b value is set to infinity which distinguishes S_j from the ordinary devices. Active devices C_i which reside in the closed proximity of S_j receive this frame and update it according to their stored information, i.e., $H_c = 1$, T_b and T_s are set according to the equation. 2 & 3 respectively. Moreover, every device C_i maintains a schedule table where valuable information about neighboring nodes is stored i.e., H_c , T_s , residual energy E_r that is calculated using equation 1.

$$E_r = E_i - E_c \quad (1)$$

where E_i and E_c represent the initial and consumed energies respectively.

Back-off timer T_b is computed using equation 2. The idea of adding an H_c value or δ with the generated random number is to minimize the collision probability of neighboring nodes as, usually, these nodes have different H_c values. However, if back-off timer T_b of the two neighboring devices are similar then these devices should recompute their T_b . δ is an infrastructure dependent parameter i.e., for flat networks its value ranges from 5-15 whereas in hierarchical networks its value ranges from 2-5.

$$T_b(C_i) = rand(0 - 1000) + min\left(\frac{T_p(C_i)}{H_c(C_i)}, \delta\right) \quad (2)$$

Transmission schedule T_s is computed using equation 3.

$$T_s(C_i) = T_b + T_p + \gamma \quad (3)$$

where T_p is the average propagation time of C_i 's first hop neighbors which includes both the transmission and processing delays. γ represents the sampling rate of a particular device which will be similar for every $C_i \in WSNs$.

Once a first hop neighboring node C_i updates the schedule-frame, it doesn't broadcast the frame immediately rather it waits for T_b . When T_b expires, C_i broadcasts an updated version of the scheduled-frame which is received by devices reside in vicinity. C_i 's neighboring devices are divided into two groups i.e., Group-I which consists of devices such that their $H_c \leq H_c(C_i)$ whereas Group-II has devices with $H_c > H_c(C_i)$. When a device $C_{i+1} \in Group - I$ receives a scheduled-frame from a neighboring device C_i it updates the schedule table entries according to the message contents and discard it. C_{i+1} discards the received scheduled-frame because it has either transmitted a scheduled-frame or waiting for its T_b to expires as it has already received a similar message from the base station module S_j . Conversely, if the scheduled-frame is received by a device $C_{i+2} \in Group - II$ then it updates the scheduled table information particularly about C_i such as H_c , T_b and T_s . Moreover, C_{i+2} computes its back-off timer using equation 2 and updates the scheduled-frame by replacing H_c , T_b and schedule time T_s with its own. When T_b of C_{i+2} expires it broadcasts the updated scheduled-frame. This process is repeatedly applied until every device $C_i \in WSNs$ in an operational network has a defined H_c value and information about neighboring node's transmission schedules T_s . Additionally, C_i 's transmission schedule is not affected even if it serves as a relaying device, that is forwarding packets of neighboring devices, in addition to its own duties.

2) *Classifier-Based Mechanism to Mitigate the Collisions Ratio of RTS/CTS and Data:* In the proposed scheme, every device C_i maintains a schedule table which contains information about neighboring devices. This information is very useful in both scenarios i.e., minimizing the collision probability and finding an optimal device.

- 1) Where multiple devices initiate a request-to-send message (RTS) at the same time and are interested to start a communication process with a shared device i.e., base station or cluster head (CH) or neighboring node.
- 2) Where a single device C_i has multiple recipient and needs to start communication with a reliable and optimal device.

In scenario-I, without a proper schedule information of neighboring devices, particularly first hop neighbors, collisions will occur and retransmission is mandatory which is not only time-consuming but power consuming too. However, if these devices are bounded to store sufficient information about neighboring devices such as T_s , T_b and H_c then packets collisions are minimized or even avoided. The proposed scheme uses a classifier-based mechanism to resolve the collision issue associated with devices interested in communication with a shared base station or other entity. Since, every neighboring device C_i has a unique back-off timer T_b , hence, the collision probability is zero even if two neighboring devices initiate the RTS process simultaneously.

In scenario-II, if a device C_i is interested to initiate a communication session with a reliable and optimal neighboring device or CH or base station then this device needs a simplified classification mechanism which identifies an optimal device. The proposed classifier-based mechanism uses various parameters such as T_s , T_b , E_r and H_c values to find an optimal neighbor. Neighboring devices are classified using equation 4.

$$C_{opt} = (W_1 * T_b + w_2 * T_s)C_i \quad (4)$$

where $W_1 = 50\%$, $W_2 = 50\%$ represent different weight-ages assigned to these parameters. A neighboring device C_i with minimum value of C_{opt} is an ideal and reliable candidate. However, if H_c and E_r of neighboring devices are not considered by our classifier then it is possible that either the transmitted packets may propagate in opposite directions or forwards to a device with minimum residual energy. In both cases the results are not favorable specifically in resource limited infrastructures, therefore, once the classifier described in equation 4 identifies the optimal neighbors then the two most optimal devices are passed to another classifier as described in equation $\xi_{reliable} = W_3 * H_c + W_4 * E_r(C_i)$ (5)

where $W_3 = 40\%$, $W_4 = 60\%$ are weight-ages assigned to the residual energy and hop-count parameters. A neighboring device C_i with maximum value of $C_{reliable}$ is considered as optimal and reliable device.

D. Implementation of the Proposed Scheme in Agricultural Environment: A Case Study

A precise and accurate DSS (preferably technology-assisted) is subjected to the selection of appropriate devices or sensors C_i , parameters to be sensed, data refinement and communication mechanisms. To accomplish this, wasp-mote



Fig. 1. Deployment of the wasp-mote agriculture boards with humidity and temperature sensors.



Fig. 2. Deployment of leaf wetness sensor in orange orchard.

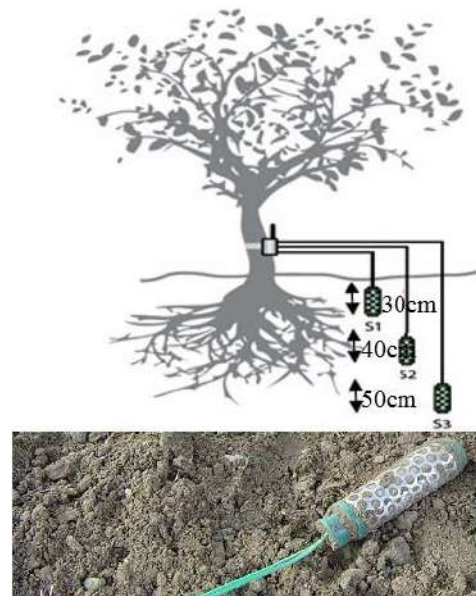


Fig. 3. Deployment of soil moisture sensor in orange orchard.

agricultural boards with a gate way were deployed in the orange orchard of our institute for approximately one year to form an automatic irrigation management system as shown in Fig. 1, Fig. 2 and Fig. 3, respectively. These boards were equipped with soil moisture, temperature, humidity and leaf wetness sensors to collect real time data continuously after a defined interval of time i.e., 30 seconds.

In the proposed DSS, soil moisture parameter is considered due to its vital role in the development of a precise watering

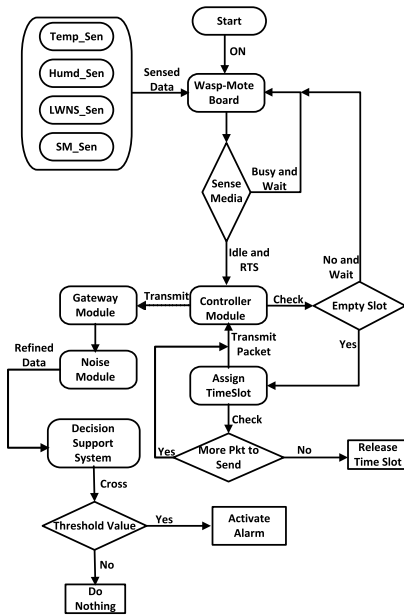


Fig. 4. Data flow diagram of the proposed WSN's based DSS for agriculture.

schedule. For example, if the sensed value is below the threshold value, then that particular area is needed to irrigated on priority basis. To further precise the proposed DSS, soil moisture sensors were deployed at three different levels in the agricultural field, as shown in Fig. 3. Likewise, atmospheric moisture exerts drastic effects on the watering schedules of various crops. Therefore, leaf wetness sensors were deployed in closed proximity to the orange leaves as shown in Fig. 2. Moreover, Temperature and humidity sensors were integrated with the wasp-mote boards to further enhance accuracy of the proposed DSS as shown in Fig. 1. The gateway module is directly connected to a computer via a USB serial port (port-6 in this case) to receive data.

In our previous data collection and communication infrastructure [6], a simplified approach was used to resolve the collision issue associated with simultaneous transmission of two or more devices C_i to a common destination. However, the system enters to deadlock if more than two devices are simultaneously transmitting to a common destination. In this paper, a modified version of the RTS/CTS handshake approach is used to resolve this issue. A detailed description of the proposed WSNs-based DSS for agriculture sector is presented in Fig. 4.

Every wasp-mote board collects real time data from various sensors i.e., temperature, humidity, soil moisture and leaf wetness. This data, say packet-x, is sent to the gateway either directly or through the relaying nodes. In both cases, the transceiver module uses the RTS/CTS handshake approach to avoid collision of packet(s). In the proposed experimental setup, the gateway module is directly connected to a central computer via USB cable specifically through port-6 and the received data is (temporarily) stored automatically using Cool Term software. Before DSS, packet-x is passed through the noise detection module to get the refined data let say packet-y. The DSS module of the proposed system checks packet-y against the threshold value, that is 250Hz for soil moisture sensor, and if the threshold value is crossed then the alarming

TABLE I
WSN'S SIMULATION PARAMETERS SETUP AND THEIR VALUES

Parameters	values
WSN Deployment Area	1000m * 1000m
Sensor Node	50, 100, 500, 1000
Base Station	One
Initial Energy (E_S)	52000 mAh
Residual Energy (E_r)	$E_S - E_c$
Packet Transmission Power Consumption (P_{T_x})	91.4 mW
Channel Delay (Ch_{delay})	10 milliseconds
Packet Receiving Power Consumption (P_{R_x})	59.1 mW
Idle Mode Power Consumption	1.27 mW
Sleep Mode Power Consumption	15.4 μ W
Transceiver Energy (T_i)	1 mW
Transmission Range (T_r)	500m
Receiving Power Threshold (RTS_n)	1024 bits
Packet Size (P_{size})	128 bytes
Initial Hop Count (H_c) of Sensor Nodes	∞
Maximum Distance between Nodes	300-450m
Sampling Rate of sensor nodes	10 to 30 seconds
Topological Infrastructure	Static and Random
Traffic Type	CBR and UDP

unit is activated along with a text message to the farmer on his mobile or LAN. If data is within the defined threshold then it is stored permanently.

IV. EXPERIMENTAL AND SIMULATION RESULTS

In this section, a detail description of both experimental and simulation results are presented to verify the exceptional performance of the proposed system against the existing schemes in terms of computational time, decisions accuracy, packet collision ratio and packet loss ratio. These algorithms were implemented in OMNET++, which is an open source simulation tool specifically designed for the resource limited networks. Initially, a static topological infrastructure, which was later on changed to the random, with a fixed propagation delay was used to mimic the real deployment process of WSNs in general and our deployment infrastructure in particular. Additionally, other networks related parameters such as interference and path-loss ratio were kept constant. A detail description of various simulation related parameters are presented in table I.

Initially, the real-time data set, that is collected through the deployment of various wasp-mote boards based tested in the orange orchard, is used as a testing tool to check the performance of these algorithms particularly in terms of computational time and decision accuracy of the underlined DSS. In terms of computational cost, the performance of these algorithms is presented in Fig. 5, which clearly depicts that the proposed algorithm performance is better than existing algorithms except the noise evading algorithm. However, a common problem associated with NE algorithm is its vulnerability to multi-valued noise, which is quite common in WSNs. Moreover, NE does not differentiate multi-valued noise from an abrupt change scenario. Similarly, the proposed scheme performance is not affected by changing data set size either statically or dynamically because it always uses a fixed sliding window. Therefore, the proposed algorithm is suitable for both scenarios, i.e., static and dynamic datasets. Additionally, these algorithms were evaluated on different static versions of the real-time dataset obtained through our deployed test bed, that

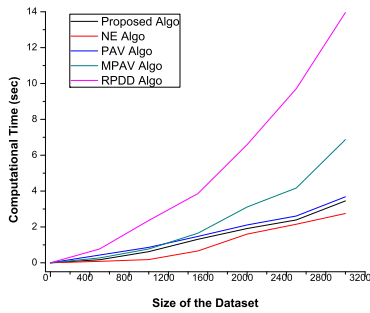


Fig. 5. Performance of DSS in terms of computational time.

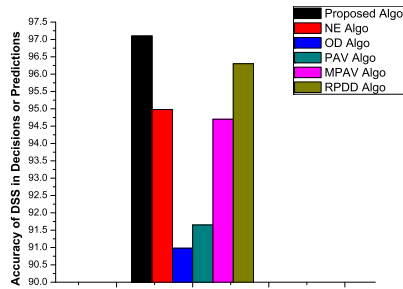


Fig. 6. Accuracy of the DSS in terms of decisions or predictions.

TABLE II
ANALYSIS OF THE PROPOSED & EXISTING ALGORITHMS ON BENCHMARK DATA SETS IN TERMS OF COMPUTATIONAL TIME

Data Set BenchMark	Proposed Algo	NE Algo [6]	PAV Algo [26]	MPAV Algo[26]	RPDD Algo[27]
50words	2.6825	2.3280	3.6090	2.8590	3.0780
B-Cancer	2.4257	2.4220	3.0780	2.7800	2.9060
Two Pattern	2.1865	2.1700	3.1880	2.7130	2.2960
Yoga	2.2958	2.1250	2.8750	2.3900	2.7030
Fish	2.3827	2.0775	2.5630	2.4850	2.5620
Mote Strain	2.8964	2.7340	3.7350	2.9370	3.0160
Diatom-Red	2.0571	2.0180	2.8600	2.0938	2.6980
Amex	2.2145	2.1090	3.4840	2.2500	3.000
Hobo Link	2.3982	2.3120	3.3590	2.4380	2.9530
Face UCR	2.6789	2.0158	3.7340	2.6400	3.4840

was approximately collected in a month or two. The proposed scheme performance is intact as shown in Fig.5.

In agriculture sector, the farmer’s attraction to the technology based infrastructures or DSS will be increased **iff** majority of their decisions or predictions are accurate. Therefore, the proposed algorithm is eager to improve accuracy of the agricultural DSS with the available computational resources and minimum cost. The decision accuracy of the proposed and existing algorithms based DSS is depicted in Fig. 6 which shows the exceptional performance of the proposed algorithm based DSS than existing algorithms. Moreover, it is evident from Fig. 6 that the NEA based DSS has a high probability of errors or wrong decision(s).

The claims of an algorithm is considered as authentic **iff** it is tested on publicly available benchmark datasets. Therefore, these algorithms were tested on various publicly available benchmark datasets as shown in Table-II. The computational time of the proposed algorithm is less than that of existing algorithms except NE which has other issues as described above. We have observed that the computational time of the proposed scheme is inversely proportional to the similarity indexes of the datasets or matching windows i.e., if similarity

TABLE III
COMPARATIVE ANALYSIS OF THE PROPOSED & EXISTING ALGORITHMS ON BENCHMARK DATA SETS IN TERMS OF ACCURACY

Data Set BenchMark	Proposed Algo	NE Algo [6]	RPDD Algo[27]	PAV Algo [26]
50words	96.21	90.40	94.66	96.59
B-Cancer	96.33	89.29	93.48	96.30
Two Pattern	96.10	89.95	93.12	95.83
Yoga	96.79	90.62	94.50	96.74
Fish	95.78	90.18	94.31	95.69
Mote Strain	95.48	89.63	93.85	95.35
Diatom-Red	96.67	90.06	94.19	96.56
Amex	96.14	87.54	93.65	95.91
Hobo Link	96.89	91.05	94.76	96.82
Face UCR	96.99	91.98	94.38	96.98

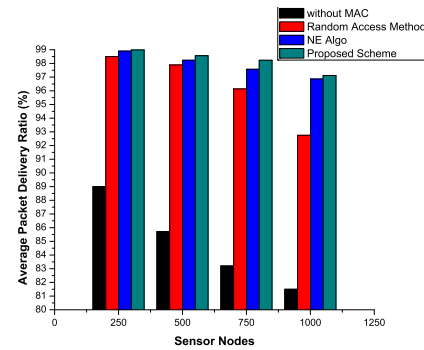


Fig. 7. Comparison of the average packet delivery ratio (simulated results).

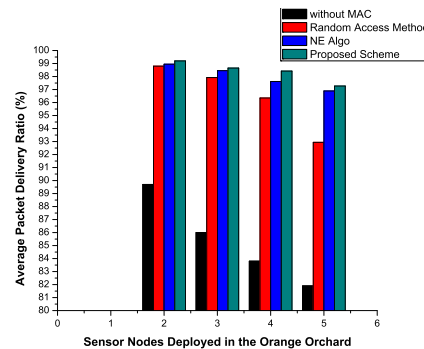


Fig. 8. Average packet delivery ratio (experimental results).

index is high the computational time will be small and vice versa. Due to the high similarity indexes of the benchmark datasets B-Cancer and Two Patterns, the computational time of the proposed algorithms is approximately equal to that of NE algorithm as shown in Table II.

Accuracy of the proposed and existing algorithms based DSS on various publicly available benchmark datasets are depicted in Table III. It is evident from Table III that the proposed mechanism is an ideal candidate for the design of an accurate and precise technology based DSS for the agriculture sector.

Packet delivery ratio is the ratio of successfully delivered packets, particularly at the destination module, to the transmitted one in an operational network. We have observed that the proposed scheme has the maximum packet delivery ratio for both real-time and simulated data against its rival schemes as shown in Fig. 7 and Fig. 8; as packet delivery ratio is inversely proportional to the packet loss ratio which is mostly due to

the packet collision. In proposed scheme, the collision issue is resolved by utilizing the RTS/CTS handshaking scheme.








V. CONCLUSION

Smart sensing-enabled networks, such as WSNs, have the ability to predict when and where the irrigation activities need to be performed. These networks enable the farmers to evaluate the required amount of water for irrigation purposes based on the data sensed by various nodes. In this paper, a real-time smart sensing-enabled Decision Support System (DSS) was presented for optimizing the water schedules for orange orchard. Smart sensing-enabled devices were deployed in different regions for approximately one year to collect soil moisture, temperature, humidity and leaf-wetness of the orchard. The gathered raw data were refined by passing it through a noise module for outliers detection. The DSS module matches the refined data against the threshold values using a modified LCSS mechanism. If these data are below the threshold value, e.g., less than 250Hz for the soil moisture sensor, then irrigation activity is scheduled in that region and the farmer is notified via a text message. Moreover, a modified version of the RTS/CTS handshake mechanism was presented to ensure the successful delivery of packets and collision avoidance. Both the experimental and simulation results showed the exceptional performance of our proposed scheme against the existing schemes for outliers detection and successful delivery of packets.

REFERENCES

- [1] G. Husak and K. Grace, "In search of a global model of cultivation: Using remote sensing to examine the characteristics and constraints of agricultural production in the developing world," *Food Secur.*, vol. 8, no. 1, pp. 167–177, Feb. 2016.
- [2] J. J. Estrada-López, A. A. Castillo-Atoche, and E. Sanchez-Sinencio, "Design and fabrication of a 3-D printed concentrating solar thermo-electric generator for energy harvesting based wireless sensor nodes," *IEEE Sensors Lett.*, vol. 3, no. 11, pp. 1–4, Nov. 2019.
- [3] H. M. Jawad *et al.*, "Accurate empirical path-loss model based on particle swarm optimization for wireless sensor networks in smart agriculture," *IEEE Sensors J.*, vol. 20, no. 1, pp. 552–561, Jan. 2020.
- [4] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, and P. Nillaor, "IoT and agriculture data analysis for smart farm," *Comput. Electron. Agricult.*, vol. 156, pp. 467–474, Jan. 2019.
- [5] A. Goap, D. Sharma, A. K. Shukla, and C. R. Krishna, "An IoT based smart irrigation management system using machine learning and open source technologies," *Comput. Electron. Agricult.*, vol. 155, pp. 41–49, Dec. 2018.
- [6] R. Khan, I. Ali, M. Zakarya, M. Ahmad, M. Imran, and M. Shoaib, "Technology-assisted decision support system for efficient water utilization: A real-time testbed for irrigation using wireless sensor networks," *IEEE Access*, vol. 6, pp. 25686–25697, 2018.
- [7] S. A. Nikolidakis, D. Kandris, D. D. Vergados, and C. Douligieris, "Energy efficient automated control of irrigation in agriculture by using wireless sensor networks," *Comput. Electron. Agricult.*, vol. 113, pp. 154–163, Apr. 2015.
- [8] B. Keswani *et al.*, "Adapting weather conditions based IoT enabled smart irrigation technique in precision agriculture mechanisms," *Neural Comput. Appl.*, vol. 31, no. S1, pp. 277–292, Jan. 2019.
- [9] D. K. Shannon, D. E. Clay, and N. R. Kitchen, *Precision Agriculture Basics*, vol. 176. Hoboken, NJ, USA: Wiley, 2020.
- [10] D. Shadrin, A. Menshchikov, D. Ermilov, and A. Somov, "Designing future precision agriculture: Detection of seeds germination using artificial intelligence on a low-power embedded system," *IEEE Sensors J.*, vol. 19, no. 23, pp. 11573–11582, Dec. 2019.
- [11] H. Panda, H. Mohapatra, and A. K. Rath, "WSN-based water channelization: An approach of smart water," in *Smart Cities—Opportunities and Challenges*. Singapore: Springer, 2020, pp. 157–166.
- [12] F. Viani, M. Bertolli, M. Salucci, and A. Polo, "Low-cost wireless monitoring and decision support for water saving in agriculture," *IEEE Sensors J.*, vol. 17, no. 13, pp. 4299–4309, Jul. 2017.
- [13] J. Gutierrez, J. F. Villa-Medina, A. Nieto-Garibay, and M. A. Porta-Gandara, "Automated irrigation system using a wireless sensor network and GPRS module," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 1, pp. 166–176, Jan. 2014.
- [14] A. Goldstein, L. Fink, A. Meitin, S. Bohadana, O. Lutenberg, and G. Ravid, "Applying machine learning on sensor data for irrigation recommendations: Revealing the agronomist's tacit knowledge," *Precis. Agricult.*, vol. 19, no. 3, pp. 421–444, Jun. 2018.
- [15] M. Dursun and S. Ozden, "A wireless application of drip irrigation automation supported by soil moisture sensors," *Sci. Res. Essays*, vol. 6, no. 7, pp. 1573–1582, 2011.
- [16] N. K. Nawandar and V. R. Satpute, "IoT based low cost and intelligent module for smart irrigation system," *Comput. Electron. Agricult.*, vol. 162, pp. 979–990, Jul. 2019.
- [17] A. Dahane, B. Kechar, Y. Meddah, and O. Benabdellah, "Automated irrigation management platform using a wireless sensor network," in *Proc. 6th Int. Conf. Internet Things, Syst., Manage. Secur. (IOTSMS)*, Oct. 2019, pp. 610–615.
- [18] T. A. Khoa, M. M. Man, T.-Y. Nguyen, V. Nguyen, and N. H. Nam, "Smart agriculture using IoT multi-sensors: A novel watering management system," *J. Sens. Actuator Netw.*, vol. 8, no. 3, p. 45, Aug. 2019.
- [19] J. Balendonck, J. Hemming, B. V. Tuijl, L. Incrocci, A. Pardossi, and A. Marzaletti, "Sensors and wireless sensor networks for irrigation management under deficit conditions (FLOW-AID)," *AgEng2008*, Crete, Greece, Tech. Rep., 2008. [Online]. Available: <https://research.wur.nl/en/publications/sensors-and-wireless-sensor-networks-for-irrigation-management-un>
- [20] L. P. Shoo *et al.*, "Moving beyond the conceptual: Specificity in regional climate change adaptation actions for biodiversity in South East Queensland, Australia," *Regional Environ. Change*, vol. 14, no. 2, pp. 435–447, Apr. 2014.
- [21] A. Pardossi *et al.*, "Root zone sensors for irrigation management in intensive agriculture," *Sensors*, vol. 9, no. 4, pp. 2809–2835, Apr. 2009.
- [22] A. N. Harun, M. R. M. Kassim, I. Mat, and S. SarahRamli, "Precision irrigation using wireless sensor network," in *Proc. Int. Conf. Smart Sensors Appl. (ICSSA)*, 2015, pp. 71–75.
- [23] O. Abrishambaf, P. Faria, L. Gomes, and Z. Vale, "Agricultural irrigation scheduling for a crop management system considering water and energy use optimization," *Energy Rep.*, vol. 6, pp. 133–139, Feb. 2020.
- [24] J. Boobalan, V. Jacintha, J. Nagarajan, K. Thangayogesh, and S. Tamilarasu, "An IoT based agriculture monitoring system," in *Proc. Int. Conf. Commun. Signal Process. (ICCCSP)*, Apr. 2018, pp. 0594–0598.
- [25] X. Dong, M. C. Vuran, and S. Irmak, "Autonomous precision agriculture through integration of wireless underground sensor networks with center pivot irrigation systems," *Ad Hoc Netw.*, vol. 11, no. 7, pp. 1975–1987, Sep. 2013.
- [26] X.-Y. Chen and Y.-Y. Zhan, "Multi-scale anomaly detection algorithm based on infrequent pattern of time series," *J. Comput. Appl. Math.*, vol. 214, no. 1, pp. 227–237, Apr. 2008.
- [27] D. T. J. Huang, Y. S. Koh, G. Dobbie, and R. Pears, "Detecting changes in rare patterns from data streams," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2014, pp. 437–448.

Service Offloading With Deep Q-Network for Digital Twinning-Empowered Internet of Vehicles in Edge Computing

Xiaolong Xu , Bowen Shen , Sheng Ding, Gautam Srivastava , Muhammad Bilal ,
 Mohammad R. Khosravi, **Varun G Menon** , Mian Ahmad Jan , and Maoli Wang 

Abstract—With the potential of implementing computing-intensive applications, edge computing is combined with digital twinning (DT)-empowered Internet of vehicles (IoV) to enhance intelligent transportation capabilities. By updating digital twins of vehicles and offloading services to edge computing devices (ECDs), the insufficiency in vehicles' computational resources can be complemented. However,

owing to the computational intensity of DT-empowered IoV, ECD would overload under excessive service requests, which deteriorates the quality of service (QoS). To address this problem, in this article, a multiuser offloading system is analyzed, where the QoS is reflected through the response time of services. Then, a service offloading (SOL) method with deep reinforcement learning, is proposed for DT-empowered IoV in edge computing. To obtain optimized offloading decisions, SOL leverages deep Q-network (DQN), which combines the value function approximation of deep learning and reinforcement learning. Eventually, experiments with comparative methods indicate that SOL is effective and adaptable in diverse environments.

Manuscript received September 6, 2020; revised October 28, 2020; accepted November 14, 2020. **Date of publication November 24, 2020;** date of current version October 27, 2021. This work was supported in part by the Financial and Science Technology Plan Project of Xinjiang Production and Construction Corps under Grant 2020DB005, in part by the NUIST Students' Platform for Innovation and Entrepreneurship Training Program under Grant 202010300024Z, and in part by the National Natural Science Foundation of China under Grant 61702277. Paper no. TII-20-4238. (Corresponding author: Maoli Wang.)

Xiaolong Xu is with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China, is with the Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China (e-mail: xlxu@ieee.org).

Bowen Shen is with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: bwshen@nuist.edu.cn).

Sheng Ding is with the Blockchain Laboratory of Agricultural Vegetables, Weifang University of Science and Technology, Weifang, 262700, China (e-mail: dingsheng@wust.edu.cn).

Gautam Srivastava is with the Department of Mathematics and Computer Science, Brandon University, Brandon, MB R7A 6A9, Canada, and also with the Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan (e-mail: srivastavag@brandou.ca).

Muhammad Bilal is with the Department of Computer and Electronics Systems Engineering, Hankuk University of Foreign Studies, Yongin-si 17035, Korea (e-mail: mbilal@kaist.ac.kr).

Mohammad R. Khosravi is with the Department of Computer Engineering, Persian Gulf University, Bushehr 75168, Iran, and also with the Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz 71557-13876, Iran (e-mail: mohammadkhosravi@acm.org).

Varun G Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala 683576, India (e-mail: varunmenon@scmsgroup.org).

Mian Ahmad Jan is with the Department of Computer Science, Abdul Wali Khan University, Mardan 23200, Pakistan (e-mail: mianjan@awkum.edu.pk).

Maoli Wang is with the School of Cyber Science and Engineering, Qufu Normal University, Qufu 273165, China (e-mail: wangml@qfnu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2020.3040180>.

Digital Object Identifier 10.1109/TII.2020.3040180

Index Terms—Deep reinforcement learning (DRL), digital twinning (DT), edge computing, Internet of vehicles (IoV), service offloading (SOL).

I. INTRODUCTION

THE INTERNET of Vehicles (IoV) is an evolution of vehicular ad hoc networks (VANETs), where vehicles are equipped with a variety of Internet of Things (IoT) equipments and envisioned as intelligent objects [1]. In the IoV, an intelligent vehicle is capable of vehicle to everything (V2X) communication. Specifically, an intelligent vehicle can share information with other vehicles through vehicle to vehicle (V2V) communications. Rather than observing the condition by a single car, V2V enables a broader view by sharing the traffic information observed by multiple vehicles, which can significantly reduce accidents caused by the blind spot [2]. Meanwhile, intelligent infrastructures like roadside units (RSUs) and smart traffic lights are deployed to analyze the vehicles in a specific region, then provide vehicles with external information through vehicle to infrastructure (V2I) communications [3]. Similarly, vehicle to pedestrian (V2P) communication enables vehicles and pedestrians to deliver commands and safety warnings [4]. With V2X communication in the IoV, intelligent vehicles have the potential to adjust the driving status in time and avoid the occurrence of traffic accidents and enhance the users driving experience.

Further, the digital twinning (DT) technology leverages machine learning and IoT technologies to create digital replicas of physical objects. The replica has its properties cloned from their original versions, and constantly update themselves with real-time data from sensors. Empowered by DT technology, a

virtual twin of vehicle in the IoV is generated and mapped to the physical vehicle with IoT technologies [5]. The DT-empowered IoV focuses on collecting the state information of the vehicle and surroundings through the smart sensor devices, and sharing the information with surrounding vehicles and infrastructures [6]. With the collected information, the digital twins are updated constantly to keep consistent with the physical vehicles. Then, through the technologies including augmented reality (AR) simulation and artificial intelligence (AI) predictive analytics, vehicles are provided with enhanced intelligence. Comparing with the traditional IoV, DT-empowered IoV can easily access the digital twins of vehicles instead of applying for and integrating numerous external data sources like the surveillance system and the remote sensing (RS) system. Under such circumstances, the data mining, simulation, and analytics of the IoV can be enhanced by DT.

As most of the collected data in the DT-empowered IoV are in the raw form (i.e., unprocessed images and videos), they cannot be directly used for control and services [7]. Thus, a powerful computing platform is required to refine the massive collected data, then feedback the extracted instructions to vehicles and passengers [8]. Usually, the processing of vehicular data requires technologies such as object detection and AR, which are computationally intensive operations [9]. To extend intelligent vehicles' capabilities, the cloud and edge computing solutions provide DT-empowered IoV with a platform as a service (PaaS) [10]. The data and service requests collected by vehicles are offloaded to the cloud data center through RSU. After data being processed at the cloud infrastructure, the refined data are fed back in the form of instructions or services [11]. Technically, the cloud data center is composed of centralized large-scale computer clusters with high performance. To reduce the cost of construction and facility maintenance, it is usually built in areas far away from end-users. Therefore, service offloading to the cloud will generate high latency during data transmission and is easy to cause bandwidth tension [12]. As a complementary paradigm of cloud computing, edge computing provides appropriate solutions in the DT-empowered IoV by offloading service requests to edge computing devices (ECDs), servers deployed close to vehicles and other end-users, for execution and data extraction [13].

Despite the advantages of fast transmission and sufficient bandwidth resources, edge computing has its own challenges. Considering the distributed manner of ECDs, the computing capacity of each independent ECD is smaller than the cloud data center. Thus, the resources in each ECD are supposed to be fully utilized to attain higher efficiency and quality of service (QoS) [14]. Further, the load balancing in ECD is an important issue, and mishandling of service offloading can cause load imbalance. Consequently, some devices in ECDs would underperform due to excessive service requests, and other would be underutilized. To enhance the performance of edge computing and provide reliable services to passengers, an effective service offloading method is needed in the DT-empowered IoV [15].

For the dynamic offloading control, deep reinforcement learning (DRL) is adopted to evaluate and choose decisions where the collective utilization is optimized [16]. Among the existing DRL

algorithms, the deep Q-network (DQN) has gained attention as a modification of Q-learning, which takes the advantage of temporal-difference learning from reinforcement learning (RL) and the function approximation from deep learning (DL) [17]. In this article, a dynamic service offloading method, named SOL, is proposed based on DQN in edge computing. Specifically, the contributions of this article are as follows.

- 1) Analyze the QoS level of DT-empowered IoV services in respect of response time in a multiuser offloading system.
- 2) Model the ECD as the agent and formalize the state, action, and reward in DRL to optimize the QoS level of the offloading system.
- 3) Apply DQN with experience replay and target network [17] to solve the problem of DT-empowered IoV service offloading in edge computing.
- 4) Conduct comparative experiments with a real-world IoV service dataset to evaluate the effectiveness and adaptability of SOL.

The rest of the article is organized as follows. In Section II, the related work is summarized. In Section III, the model of service offloading in edge computing is described. In Section IV, details of DRL and SOL are presented. Then, in Section V, comparison experiments are conducted. Finally, Section VI concludes this article.

II. RELATED WORK

So far, various applications in the DT-empowered IoV have been proposed to enhance the QoS, safety, and security of transportation [18]. However, the generated data of such applications are large in scale and has much redundancies, therefore not suitable for local computing and existing cloud computing paradigms [19]. Hu *et al.* [20] addressed the scale-sensitive problem of existing object detection, then modified the deep convolutional neural network for vehicle detection with a large variance of scales to guarantee the accuracy and safety in IoV. From another perspective, Liu *et al.* [21] exhibited the outstanding performance of edge computing on enhancing the security and QoS of autonomous vehicles, including extending computing capacity and reducing energy consumption.

The placement of ECDs has great impact on overall performance of edge computing. Zhao *et al.* [22] proposed a ranking-based near-optimal placement algorithm to minimize average access delay through SDN techniques in cloudlets placement. Wang *et al.* [23] studied the ES placement while considering load balancing as well as access delay and adopted mixed integer programming to find the optimal placement. After ECDs are located, task offloading can be taken into operation. He *et al.* [24] gave consideration to users' privacy and system cost in mobile edge computing, and proposed a novel task offloading scheme to enhance user experience. Zhou *et al.* [25] investigated the task offloading under information asymmetry and uncertainty in vehicular fog computing, and proposed a contract optimization to realize the effective server recruitment.

Owing to higher effectiveness of evolutionary algorithms (EAs), researchers widely adopted EAs as a tool for optimizing the offloading problems in edge computing. Guo *et al.* [26]

TABLE I
NOTATIONS AND DEFINITIONS

Notations	Definitions
N	The number of RSUs
M	The number of ECDs
K	The number of vehicles
R	The set of RSUs, $R = \{r_1, r_2, \dots, r_N\}$
E	The set of ECDs, $E = \{e_1, e_2, \dots, e_M\}$
V	The set of vehicles, $V = \{v_1, v_2, \dots, v_K\}$
$D(t)$	The data size of services at time t , $D(t) = \{d_1(t), d_2(t), \dots, d_K(t)\}$
C_e	The coverage of ECD
C_r	The coverage of RSU
$dist$	The distance between two network nodes
RT	The response time of services
S	The QoS level of services

comprehensively investigated the computation offloading as a mix integer nonlinear programming problem, and designed a computation algorithm based on the genetic algorithm and particle swarm optimization to minimize the energy consumption of the user equipment. However, EAs are usually iterative algorithms that find the global optimal solutions by updating the current solutions continuously. Thus, the dependency on global information and the considerable time complexity during the iteration of generations become significant drawbacks [27]. If EAs are adopted for the offloading of each service, the time overhead in controlling can be unaffordable for the practical implementation of edge computing-empowered IoV.

To obtain decentralized and time-efficient control in the IoV, DRL has been adopted in many aspects of the IoV. To achieve high QoS V2V communication, a decentralized resource allocation mechanism based on DRL is designed in [28]. Benefitting from the decentralized manner, DRL can significantly reduce the transmission overhead and the waiting time for global information. Apart from the efficiency, DRL also exhibits the advantage in adaptability. Liang *et al.* [29] adopted DRL to study the automatic determination of traffic signal duration based on the data collected from sensors. In their model, the actions are changes in the duration of a traffic light, and the reward is the difference in cumulative waiting time between two signal cycles. Meanwhile, Zhou *et al.* [30] proposed a DRL-based car-following model, which can make adjustments in driving behaviors under diverse traffic demands, to improve travel efficiency and safety at signalized intersections in real-time. Generally, DRL is promising in achieving distributed control in the dynamic environment of IoV.

III. SYSTEM MODEL AND PROBLEM DEFINITION

This section describes the system model and service offloading in edge computing. Table I presents the key notations and definitions used in this article.

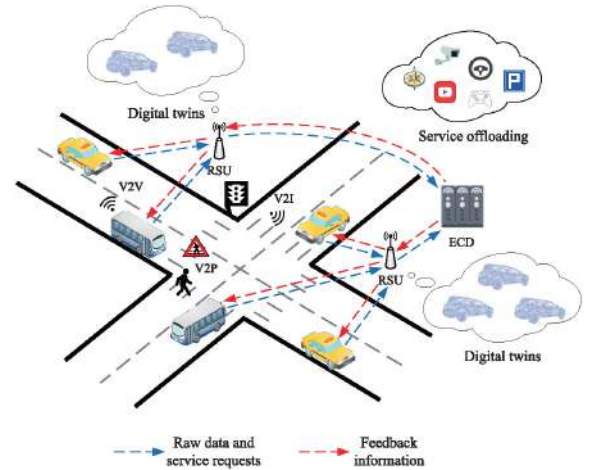


Fig. 1. Framework of service offloading in DT-empowered IoV with edge computing.

A. Framework of Service Offloading for DT-Empowered IoV in Edge Computing

In the proposed framework, vehicles are denoted by set $V = \{v_1, v_2, \dots, v_K\}$. For each vehicle, a digital twin of itself is generated with information of position, speed, vehicle gap, and dashcam videos collected by vehicular sensors and cameras. The raw data and service messages of vehicles can be sent to RSUs, denoted by set $R = \{r_1, r_2, \dots, r_N\}$. With the constant update, we can assume that the cloning is successful, and the functions of the digital twin keep pace with the entity's. Each vehicle can concurrently request one service at time t , and the data to be processed of each vehicular service is denoted by set $D(t) = \{d_1(t), d_2(t), \dots, d_K(t)\}$, while $d_i(t) = 0$ indicates that no service is requested by vehicle v_i . For RSUs are usually considered as communicating nodes and not capable of a large scale of computing tasks, ECDs are arranged to some certain districts to process the service requests based on digital twins of vehicles with massive data collected by RSUs. The ECDs are denoted by the set $E = \{e_1, e_2, \dots, e_M\}$. RSUs can communicate with each other as well as ECDs in their transmission range. Generally, the framework of task offloading in DT-empowered IoV with edge computing is shown in Fig. 1.

In the DT-empowered IoV, the coverage of each ECD is assumed to be the same and denoted by C_e , while for RSUs, the range is denoted by C_r . Then, every RSU, ECD, and vehicle can be, respectively, denoted by

$$r_i(\text{lat}_i, \text{lon}_i, C_r), \quad 1 \leq i \leq N \quad (1)$$

$$e_j(\text{lat}_j, \text{lon}_j, C_e), \quad 1 \leq j \leq M \quad (2)$$

$$v_k(\tilde{\text{lat}}_k(t), \tilde{\text{lon}}_k(t), d_k(t)), \quad 1 \leq k \leq K \quad (3)$$

where lat_n and lon_n represent the latitude and longitude of a network node, respectively, as the location of vehicle is dynamic with time, $(\tilde{\text{lat}}_k(t), \tilde{\text{lon}}_k(t), d_k(t))$ is used to represent the state of v_k at time t .

Based on the latitude and longitude, the distance between two nodes (i.e., RSU, ECD, or cloud access point) can be calculated by the Euclidean distance as

$$\text{dist}(\text{node}_i, \text{node}_j) = \sqrt{(\text{lat}_i - \text{lat}_j)^2 + (\text{lon}_i - \text{lon}_j)^2}. \quad (4)$$

It is guaranteed that the data transmission between a vehicle and an RSU, as well as an RSU and an ECD, is a one-hop transmission. Specifically, each RSU is in the coverage of at least one ECD while each vehicle is in the coverage of at least one RSU as

$$\forall r_i \in R, \min_{e_j \in E} \text{dist}(r_i, e_j) \leq C_e \quad (5)$$

$$\forall v_k \in V, \min_{r_i \in R} \text{dist}(v_k(t), r_i) \leq C_r. \quad (6)$$

B. QoS Model of DT-Empowered IoV Services Offloading in Edge Computing

RSUs in the offloading system can independently choose their computing paradigm in each time period, namely, local computing or edge computing. The response time of a service request can be calculated as the sum of offloading time, execution time, and feedback time.

1) Local Computing Model: When vehicle v_k proposes a service request at time t , and locally executes it, the offloading indicator is $a_k(t) = 0$. In this case, local computing yields a response time of $RT_k^l(t)$, which only includes the execution time of the task by vehicular computing units. The execution time is determined by the processing capacity of resource units and the length of data to be executed. Considering that the processing requirements of vehicular services are usually different, a standard measurement is to divide the vehicular processor into multiple resource units with same local computing capacity of λ_l^{exec} , and u_a of these units are activated for the service. Then the local execution time is calculated as

$$RT_k^l(t) = RT_k^{\text{que}}(t) + \frac{f(d_k(t))}{u_a \cdot \lambda_l^{\text{exec}}} \quad (7)$$

where $RT_k^{\text{que}}(t)$ is the queuing time of the task, denoted by the difference between the execution starting time and requested time as $RT_k^{\text{que}}(t) = T_k^{\text{start}} - T_k^{\text{request}}$. Meanwhile, $f(d_k(t))$ represents the total computation of the service with the size $d_k(t)$ of raw data.

2) Offloading Computing Model: When the service of vehicle v_k is determined to be offloaded to ECD, the offloading indicator is $1 \leq a_k(t) \leq M$, which indicates that the offloading destination is the $a_k(t)$ th ECD in the offloading system. Accordingly, the response time $RT_k^o(t)$ is generated during three parts of offloading computing. First, the data and service request of vehicle are transmitted from v_k to the nearest RSU r_i , and r_i offloads the service to the destination ECD. During this phase, network latency occurs in the data transmission, calculated as

$$\begin{aligned} RT_k^{o,\text{tran}}(t) &= RT_v^{o,\text{tran}}(t) + RT_r^{o,\text{tran}}(t) \\ &= \frac{d_k(t)}{\lambda_v^{\text{tran}}} + \frac{d_k(t)}{\lambda_r^{\text{tran}}} \end{aligned} \quad (8)$$

where λ_v^{tran} is the data transmission rate between v_k and r_i while λ_r^{tran} are the data transmission rate between r_i and ECD. According to the Shannon–Hartley theorem, λ_v^{tran} and λ_r^{tran} is affected by the bandwidth B of the channel, signal power p_t , and the average power of the additive white Gaussian noise p_n . As the channel resources of an RSU are often utilized by several vehicles, the bandwidth utilized by each RSU is denoted by $\frac{B}{K_c}$ when K_c vehicles are utilizing the channel concurrently. Thus, λ_v^{tran} is calculated as

$$\lambda_v^{\text{tran}} = \frac{B_r}{K_c} \log_2 \left(1 + \frac{p_t}{p_n} \right) \quad (9)$$

analogously, the transmission rate λ_r^{tran} between the ECD and one of N_c RSUs is calculated as

$$\lambda_r^{\text{tran}} = \frac{B_e}{N_c} \log_2 \left(1 + \frac{p'_t}{p'_n} \right). \quad (10)$$

After the service and digital twin data of v_k being offloaded, the destination ECD will take time for execution. Analogous to (7), the execution time of ECD is calculated as

$$RT_k^{o,\text{exec}}(t) = RT_k^{\text{que}}(t) + \frac{f(d_k(t))}{u_a \cdot \lambda_o^{\text{exec}}} \quad (11)$$

where λ_o^{exec} represents the execution capacity of the ECD, usually considered as $\lambda_o^{\text{exec}} = n \cdot \lambda_l^{\text{exec}}$.

After the task is executed, the computing results are reported back to the RSU to update the digital twin and give instruction to the vehicle. Usually, the feedback data are condensed with a relatively small size of d'_k . Thus, the feedback time during feedback is considered negligible.

Based on (8) and (11), the total response time of the service proposed by v_k at time t by offloading computing is $RT_k^o(t) = RT_k^{o,\text{tran}}(t) + RT_k^{o,\text{exec}}(t)$.

3) QoS Measurement: To quantify and measure the QoS, the maximum tolerable response time RT_{th} is used as a standard to normalize the indicator of QoS. The QoS level of response time in local computing and offloading computing are calculated as

$$S_k^l(t) = 1 - \frac{RT_k^l(t)}{RT_{th}} \quad (12)$$

$$S_k^o(t) = 1 - \frac{RT_k^o(t)}{RT_{th}}. \quad (13)$$

C. Problem Definition

In the multiuser offloading system, the goal is to maximize the average QoS level of vehicular services through an optimal offloading strategies set $A(t) = \{a_1(t), a_2(t), \dots, a_K(t)\}$ at each time period t . Based on the models given above, the problem of service offloading in DT-empowered IoV is formulated as

$$\max_{A(t)} \sum_{k=1}^K \left[S_k^l(t) + \sum_{m=1}^M S_k^o(t) \Pr[a_k(t) = m] \right] / \sum_{k=1}^K \text{Sgn}(d_i(t)) \quad (14)$$

$$s.t. \quad \forall v_k \in V, a_k(t) \in [0, M] \quad (15)$$

$$\forall v_k \in V, S_k^o(t) \geq 0, S_k^l(t) \geq 0 \quad (16)$$

where $\Pr[a_k(t) = m]$ is the probability of $a_k(t) = m$, i.e., the value is 1 if $a_k(t) = m$, otherwise, 0. Meanwhile, $\text{Sgn}(d_i(t))$ is the sign of $d_i(t)$, i.e., $\text{Sgn}(d_i(t)) = 1$ indicates that $d_i(t)$ is positive, and when $d_i(t) = 0$, $\text{Sgn}(d_i(t)) = 0$. As an element of A , $a_k(t)$ represents the offloading destination, subject to constraint (15). When $a_k(t) = 0$, the service will be locally executed. Otherwise, it will be offloaded to the corresponding ECD for execution. Meanwhile, equation (16) indicates that the QoS is not negative, i.e., the service response time should be within the maximum tolerable time.

IV. SOL FOR DT-EMPOWERED IOV SERVICES OFFLOADING

In this section, SOL is designed for the service offloading in edge computing-enabled IoV. First, the framework of RL is introduced in service offloading. Then, the drawback of a primitive RL algorithm-named Q-learning is analyzed, and a DRL algorithm named DQN is leveraged for SOL.

A. Framework of Reinforcement Learning in SOL

RL is one of the significant branches of machine learning alongside supervised learning and unsupervised learning. It refers to the process of achieving the highest cumulative rewards through the exploration of the environment and the exploitation of previous knowledge. During such a trial-and-error process, the agent in RL can obtain the perception of the environment and the decision-making strategy.

In the offloading system, the ECD is enabled the controlling of offloading decisions and viewed as the agent in RL. There are three key elements of an agent, namely, the state (s), the action (a), and the reward (R). Usually, the state is also considered the environment that the agent reacts to. In SOL, the state consists of two components, the available units of ECD, and the average QoS level of each vehicle in the offloading system calculated as (14). When the ECD receives a service request, it searches for an optimal action $a_k(t)$ available in its current state. Based on the action indicator $a_k(t)$, the ECD decides where to offload and execute the service request. After making offloading decision and execution, the QoS level of service is evaluated in terms of the vehicle's response time as $S_k(t)$, then fed back to ECD as the reward. In general, the goal of RL is to obtain the highest cumulative reward in a learning episode.

Among the RL algorithms, Q-learning has proved to be effective in model-free learning problems [31]. In Q-learning, the agent is given a Q-table which records the Q-value (i.e., quality) of each pair of state and action as $Q(s, a)$. For each step, the agent selects an action a_t at the state s_t which brings it the highest reward, then calculates and updates $Q(s_t, a_t)$ based on the action it chooses and the reward it gets as

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \delta_t \quad (17)$$

where α is the learning rate parameter that satisfies $0 \leq \alpha \leq 1$ and determines the extent to which the newly acquired knowledge overrides the old knowledge. Meanwhile, δ_t is the difference between the actual value of $Q(s_t, a_t)$ and the estimated

value of it through the Q-table, calculated as

$$\delta_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \quad (18)$$

where γ represents the discount factor of future reward, s_{t+1} is the next state after the agent performing a_t , and r_t is the instant reward experienced by the agent, also denoted as the QoS level of service. Notice that, if the response time exceeds the maximum tolerable time, the reward r_t is set as $r_t \leftarrow \min(r_t, 0)$ automatically as a punishment. Specifically, the discount factor satisfies $0 \leq \gamma \leq 1$, and the larger γ means that the agent has a clearer view toward the future while lower γ means that the agent is more focused on the instant reward. Usually, Q-learning starts with a lower discount factor and increases it toward its final value to accelerate learning.

As directly choosing the action with maximal Q-value encourages exploitation but lacks exploration, agents might fall into the local optimum. Thus, a certain degree of randomness is allowed by introducing the ϵ -greedy in strategy selection. Specifically, agents select the strategy with the highest Q-value with probability $\Pr[s_i(t) = s_{\text{best}}] = 1 - \epsilon$ to exploit knowledge, while with probability ϵ , they randomly select another action to explore for more available choice. Usually, ϵ decreases over time to encourage exploration during the early phase and limit the blindness and fluctuation of agents' decision-making in the later phase.

B. SOL With Deep Q-Network

The primitive reinforcement learning method has a significant disadvantage that it requires a Q-table to store the Q-values of all possible state-action pairs. However, the number of states is large or even infinite, the traverse and update of Q-table become time-consuming. Moreover, there exist many state-action pairs that are similar but not identical in a complex Q-table. Therefore, the traditional Q-learning method will become ineffective since the possibility of the agent to access a specific state-action pair is relatively small. To tackle the problem, a practical approach is to approximate the Q-values of different state-action pairs with deep neural network (DNN), which leads to the primary essence of DQN [17]. Intuitively, the differences between Q-learning and DQN in offloading decision-making are shown in Fig. 2.

Practically, the proposal of DQN successfully combined RL with DL while tackling the challenges in the inconsistency between them. Usually, DL assumes that the distribution of data samples is in an independent manner. However, the states and actions in RL are usually highly correlated, which is not consistent with the requirement of DL. To mitigate the correlation in data, a technique of experience replay is introduced. Technically, a structure of experience pool D , which stores the experience of each step as $e_t(s_t, a_t, r_t, s_{t+1})$, is adopted to enable experience replay in DQN. During the network training, a minibatch of the experience is randomly drawn from D for training, such that the distribution of data can be averaged, and the correlations can be alleviated.

Another feature of DQN is to generate a target Q value in a separate network (i.e., the target network Q^{tar}). Unlike the original network (i.e., the prediction network Q^{pre}) which updates

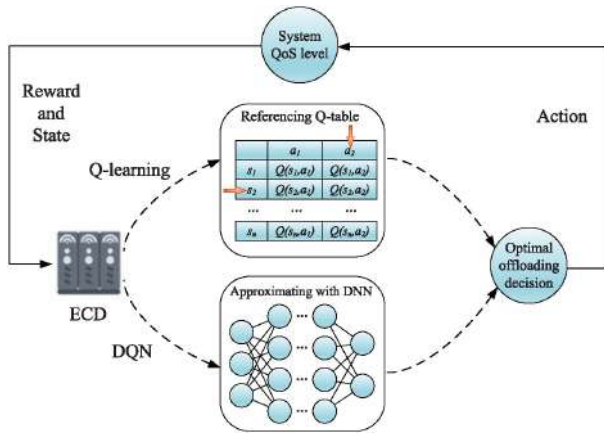


Fig. 2. Differences between offloading decision-making based on Q-learning and DQN.

the parameters θ in every iteration, θ^- in the target network are only periodically updated in every C iterations and stay fixed in other steps. Specifically, after C rounds of updates by the prediction network, the target network is updated by a copy of the prediction network. This feature adds a delay between the update of the network and the effect on the targets y_j and further stabilizes the performance of DQN.

With DNN, the Q-value of state-action pair (s_t, a) is estimated as $Q^{\text{pre}}(s_t, a; \theta) \approx Q(s_t, a)$, where the parameter θ is a vector of weights in the DNN. To evaluate the accuracy of the approximation and further train the network, the loss function is introduced as

$$L_i(\theta_i) = \mathbb{E} \left[(y_i - Q^{\text{pre}}(s, a; \theta_i))^2 \right] \quad (19)$$

where y_i represents the target Q-value generated by the target network of

$$y_i = r + \gamma \max_{a'} Q^{\text{tar}}(s_{t+1}, a', \theta_i^-). \quad (20)$$

By minimizing $L_i(\theta_i)$ through updating weight θ repeatedly, the network can be trained to be more accurate. Technically, minibatch stochastic gradient descent (MSGD) is applied to minimize the difference between the output of the target network and the prediction network. More precisely, the pseudo code of DQN is shown in Algorithm 1.

C. SOL Review

Generally, SOL is designed on the logical basis shown in Fig. 3. The basic idea of SOL is to enable the ECD to make optimal offloading decisions through RL. With the exploration of the unknown environment, the agent in RL can learn from the feedback reward. Meanwhile, the exploitation of experienced knowledge enables the agent to select optimal action at each state, jointly considering the instant reward and long-term reward. However, as the environment of the IoV service offloading system is dynamic and sophisticated, the space of states can be vast or infinite. If primitive RL algorithms like Q-learning are adopted, the update and search for optimal offloading decisions generate a significant overhead of storage and time. Moreover,

Algorithm 1: SOL With Deep Q-Network.

- 1: Initialize experience pool D with the size of N
- 2: Initialize Q^{pre} and Q^{tar} with same random weights θ
- 3: **for** episode = 1 to M **do**
- 4: **for** $t = 1$ to T **do**
- 5: Approximate Q-values of all actions at state s
- 6: Select the optimal offloading decision a_t based on ϵ -greedy policy
- 7: Perform service offloading or local computing according to a_t
- 8: Calculate the reward r_t and the next state s_{t+1}
- 9: Store experience $e_t(s_t, a_t, r_t, s_{t+1})$ in D
- 10: Perform MSGD to update the parameters θ of prediction network Q^{pre} through minimizing $L(\theta)$
- 11: Update the target network Q^{tar} every C steps
- 12: **end for**
- 13: **end for**

the similar but not identical states significantly increase the agent's exploration range and will lead to slow convergence of RL. To reduce the overhead in storage and time while fastening convergence, a DRL algorithm named DQN is adopted in SOL. Instead of referencing the Q-table to find the optimal decision, DQN introduced the function value approximation of DL to estimate the Q-value of state-action pairs. Also, with the features of experience replay and target network, DQN successfully alleviates the inconsistency between RL and DL, and can achieve satisfying performance in SOL.

V. EXPERIMENTAL EVALUATION

In this section, SOL is implemented and experiments are conducted based on the real-world IoV service requests. Then, comparative offloading strategies are introduced. Finally, the results of SOL and comparative offloading strategies under different circumstances are presented, and the effectiveness and adaptability of SOL are verified based on the experimental results.

A. Experiment Setup

Two real datasets of IoV service requests in Nanjing are applied in the experiment. One dataset contains details of 436 activated RSUs in Nanjing, including their latitude and longitude values. Based on the RSU locations, partitioning around medoids (PAM) clustering is adopted with the parameter $K=40$ to simulate the placement of ECDs and the assignment of RSUs. As shown in Fig. 4, on part of the brief road map of Nanjing, the RSUs and ECDs in one cell of the offloading system are marked with blue dots and red server icons, respectively. The 3 ECDs and 26 RSUs (including 3 colocated with each ECD) are analyzed in the experiments.

The other dataset contains vehicular service requests collected by RSUs in 30 consecutive days (from 00:00:00 Sep. 1st to 23:59:59 Sep. 29th). The total number of service requests is more than 160 million. From the second dataset, the service

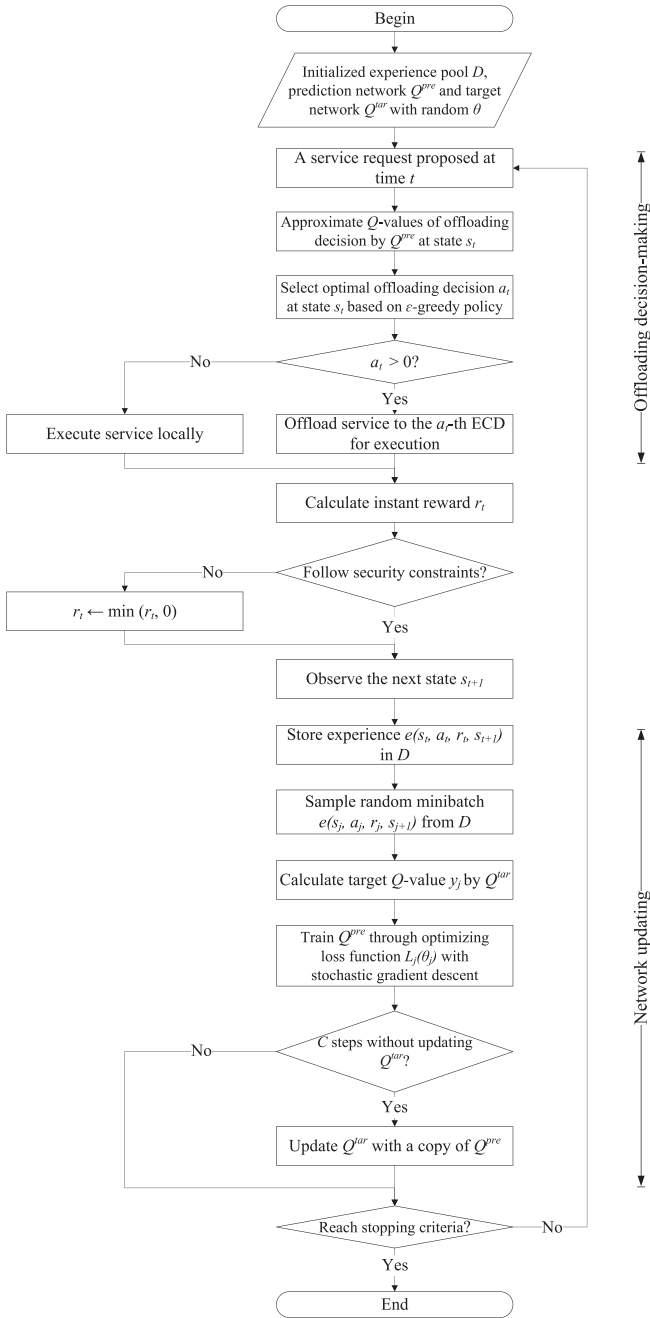


Fig. 3. Programming flowchart of SOL.

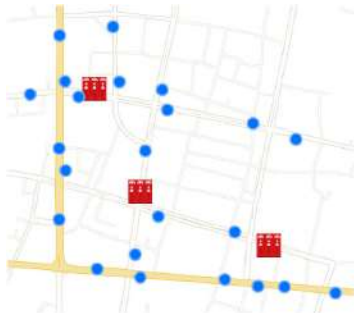


Fig. 4. Distribution of RSUs and ECDs in an offloading system.

TABLE II
CONTROLLED VARIABLE SETTINGS

Variable description	Controlled value
ECD execution capacity	$5 \times$ local execution capacity
Number of ECD	3
Number of service requests	5 per vehicle
Average size of raw data	50 MiB per request

requests in one cell of the offloading system are extracted for comparative analysis.

B. Comparative Offloading Strategies

1) *Entirely Local Computing*: Entirely local computing is a conventional paradigm which depends only on the vehicles' local execution capacity. Entirely local computing requires no additional controlling strategy, and is used as a baseline to evaluate the optimization capability of other offloading strategies.

2) *Nearest Neighbor Offloading Computing*: Contrary to entirely local computing, nearest neighbor offloading strategy enables all the service requests and raw data to be offloaded to the nearest ECD for execution. As the location of RSUs and ECDs are fixed, the nearest ECD of each RSU can be determined. When the computational resources of ECD are abundant, this strategy can achieve a high level of QoS without complicated controlling. However, as the local computing units are not utilized, and the distribution of workload is uneven, excessive offloaded services will increase the risk of ECD being overloaded and severely lower the QoS level of the offloading system.

3) *First Fit Offloading Computing*: First fit is an online algorithm where the service is offloaded to the nearest ECD that can accommodate it. When first fit algorithm begins, it searches for the closest ECD to the RSU which collected a service request. If the ECD has insufficient idle resource units for the service, it will be offloaded to the next closest ECD with sufficient resources. If no ECD is capable, the service will be executed by the computing devices of the vehicle which proposes the request.

C. Analysis on the Adaptability of Offloading Strategy

As the real condition of IoV services in cities are various, e.g., the number of vehicles and ECDs varies with the development of cities. Thus, the offloading strategy needs to be adaptive, so that it can be applied widely. To verify the adaptability of SOL, four sets of controlled experiments with diversity in services conditions are conducted, and the performance of SOL is evaluated.

The controlled value of variables in the comparative analysis are listed in Table II. In each set of experiment, there is one variable with its value fluctuating around the controlled value and the others remain unchanged.

1) *Analysis on the Variety of ECD Execution Capacity*: Experiments are conducted with different ECD's execution capacity, and the results are shown in Fig. 5. In this set of experiments, the ratio of ECD execution capacity to local execution capacity

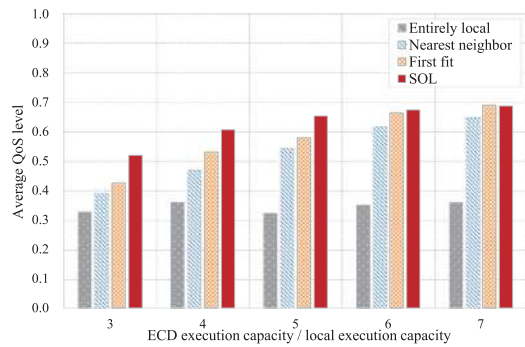


Fig. 5. Comparison of QoS level with variety in ECD capacity.

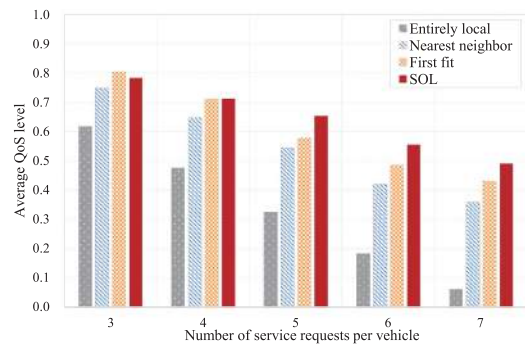


Fig. 7. Comparison of QoS level with variety in service number.

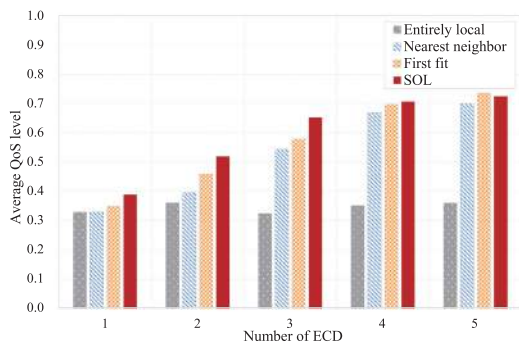


Fig. 6. Comparison of QoS level with variety in ECD number.

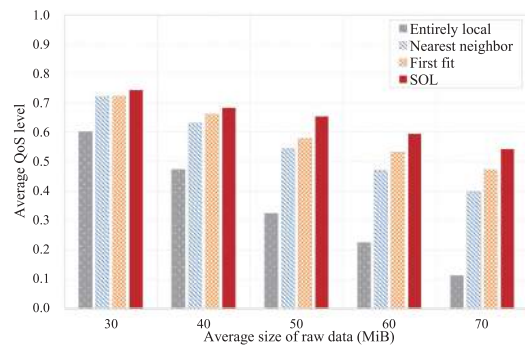


Fig. 8. Comparison of QoS level with variety in average size of raw data.

ranges from three to seven. As the results indicate, SOL outperforms entirely local computing, nearest neighbor offloading, and first fit offloading in response time. When the capacity of ECD is insufficient, the risk of ECD being overloaded is high if no effective offloading strategy is adopted. Thus, the QoS level of vehicular services by nearest neighbor offloading is severely reduced by long response time. In contrast, when the execution capacity of ECD is ample, the difference in response time between SOL and the other offloading strategies is small. As the ECDs can efficiently execute most of the services, offloading computing is usually the optimal choice.

2) Analysis on the Variety of ECD Number: When the number of ECDs in the offloading system are different, the QoS level of vehicular services are shown in Fig. 6. With other variables unchanged, the number of ECDs ranges from one to five in this set of experiments. The QoS level of SOL is generally the highest despite the little disadvantage over first fit when ECD number is five. When ECDs are sparsely deployed, the ECDs can be easily overloaded by the excessive service requests. Thus, SOL tends to assign the services to be executed locally and has a slight advantage over other strategies. In contrast, when ECDs are ample, the service requests and the workload of ECDs are more balanced with SOL or first fit offloading strategy, and overloading is unlikely to occur during offloading computing.

3) Analysis on the Variety of Service Number per Vehicle: Fig. 7 illustrates the impact on the QoS level by the number of services per vehicle. In this set of experiments, we assume each vehicle can propose multiple service requests at different time, and the number of proposed service requests per vehicle ranges

from three to seven, while other variables remain unchanged. The QoS level of response time by offloading method goes down as the number of services rises, while SOL keeps the decline smaller than first fit and nearest neighbor offloading. The advantage of SOL is that ECD selectively executes some of the services while others are executed locally, which reduces the latency in queuing. When the execution capacity of ECD goes beyond the service requests of vehicles, the QoS level of first fit and nearest neighbor offloading is close to the one of SOL and both outperform local computing. In addition, as the bandwidth of ECD is usually considered fixed, the intensive data transmission also has an impact on the offloading time when the communication is frequent.

4) Analysis on the Variety of Average Size of Raw Data: In Fig. 8, the QoS level with diversity in the average size of raw data is analyzed. Experiments are conducted with the average size of raw data ranging from 30 to 70 MiB, while the other variables remain unchanged. It is intuitive that the QoS level declines with the rise in the size of raw data. As the computing capacity of on-board devices is usually insufficient, the response time of local computing is intolerable. Simultaneously, the QoS level of nearest neighbor offloading, first fit offloading, and SOL also experience a drop. However, as the execution rate of ECD is much higher than on-board devices, the increase in response time by offloading methods is not significant. Instead, the time overhead generated in data transmission has an impact on the QoS level. Hopefully, 5G communication is promising in mitigating the data transmission time and further enhance the QoS level of service offloading by SOL.

VI. CONCLUSION

In this article, edge computing was adopted in the DT-empowered IoV to provide vehicular services with a high QoS level, and a service offloading method with deep reinforcement learning named SOL is proposed. First, a multiuser offloading system in DT-empowered IoV was modeled with consideration of response time. Then, DQN with experience replay and target network, which exerts the advantages of both RL and DL, was adopted in the offloading system to obtain optimal offloading strategy. The experiments were conducted with a real-world dataset of RSU locations and IoV service requests, and the results verified the effectiveness and adaptability of SOL.

To simplify the model, the IoV service offloading was modeled as a binary offloading process where the services are assumed atomic, i.e., services cannot be divided and executed on more than one devices. In future works, partial offloading can be taken into consideration where a service can be divided into several procedures and offloaded to different ECDs. In this case, computational resources can be better utilized. However, if partial offloading is adopted, the partibility, dependency, and priority in the procedures of services need to be thoroughly analyzed, and the offloading decisions are required a strict graph dependency constraint.

REFERENCES

- [1] J. Contreras-Castillo, S. Zeadally, and J. A. Guerrero-Ibañez, "Internet of Vehicles: Architecture, protocols, and security," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3701–3709, Oct. 2018.
- [2] K. Asano, N. Enami, T. Kamada, and C. Ohta, "Person reidentification for detection of pedestrians in blind spots through V2V communications," in *Proc. IEEE Intell. Transport Syst. Conf.*, 2019, pp. 764–770.
- [3] L. Chen and C. Englund, "Cooperative intersection management: A survey," *IEEE Trans. Intell. Transport Syst.*, vol. 17, no. 2, pp. 570–586, Feb. 2016.
- [4] X. Wang *et al.*, "Optimizing content dissemination for real-time traffic management in large-scale Internet of Vehicle systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1093–1105, Feb. 2019.
- [5] O. Veledar, V. Damjanovic-Behrendt, and G. Macher, "Digital twins for dependability improvement of autonomous driving," in *Systems, Software and Services Process Improvement*, Cham, Switzerland: Springer, 2019, pp. 415–426.
- [6] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models," *Proc. IEEE*, vol. 108, no. 10, pp. 1785–1824, Oct. 2020.
- [7] X. Wang, L. T. Yang, L. Song, H. Wang, L. Ren, and J. Deen, "A tensor-based multi-attributes visual feature recognition method for industrial intelligence," *IEEE Trans. Ind. Informat.*, early access, doi: [10.1109/TII.2020.2999901](https://doi.org/10.1109/TII.2020.2999901).
- [8] M. Zhang, C. Chen, T. Wo, T. Xie, M. Z. A. Bhuiyan, and X. Lin, "SafeDrive: Online driving anomaly detection from large-scale vehicle data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2087–2096, Aug. 2017.
- [9] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big data analytics in intelligent transportation systems: A survey," *IEEE Trans. Intell. Transport Syst.*, vol. 20, no. 1, pp. 383–398, Jan. 2019.
- [10] K. Djemame *et al.*, "PaaS-IaaS inter-layer adaptation in an energy-aware cloud environment," *IEEE Trans. Sustain. Comput.*, vol. 2, no. 2, pp. 127–139, Apr./Jun. 2017.
- [11] M. Abbasi, M. Rafiee, M. R. Khosravi, A. Jolfaei, V. G. Menon, and J. M. Koushyar, "An efficient parallel genetic algorithm solution for vehicle routing problem in cloud implementation of the intelligent transportation systems," *J. Cloud Comput.*, vol. 9, no. 1, 2020, Art. no. 6.
- [12] V. G. Menon, S. Jacob, S. Joseph, and A. O. Almagrabi, "SDN-powered humanoid with edge computing for assisting paralyzed patients," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5874–5881, Jul. 2020.
- [13] J. Ren, G. Yu, Y. He, and G. Y. Li, "Collaborative cloud and edge computing for latency minimization," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5031–5044, May 2019.
- [14] Q. He *et al.*, "A game-theoretical approach for user allocation in edge computing environment," *IEEE Trans. Parallel Distrib. Syst.*, vol. 31, no. 3, pp. 515–529, Mar. 2020.
- [15] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7944–7956, Aug. 2019.
- [16] Z. Ning *et al.*, "When deep reinforcement learning meets 5G vehicular networks: A distributed offloading framework for traffic big data," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1352–1361, Feb. 2020.
- [17] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [18] Z. Wang *et al.*, "A digital twin paradigm: Vehicle-to-cloud based advanced driver assistance systems," in *Proc. IEEE 91st Veh. Technol. Conf.*, 2020, pp. 1–6.
- [19] X. Wang, L. T. Yang, Y. Wang, L. Ren, and M. J. Deen, "ADTT: A highly-efficient distributed tensor-train decomposition method for iiot big data," *IEEE Trans. Ind. Informat.*, early access, doi: [10.1109/TII.2020.2967768](https://doi.org/10.1109/TII.2020.2967768).
- [20] X. Hu *et al.*, "Sinet: A scale-insensitive convolutional neural network for fast vehicle detection," *IEEE Trans. Intell. Transport Syst.*, vol. 20, no. 3, pp. 1010–1019, Mar. 2019.
- [21] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1697–1716, Aug. 2019.
- [22] L. Zhao, W. Sun, Y. Shi, and J. Liu, "Optimal placement of cloudlets for access delay minimization in SDN-based Internet of Things networks," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1334–1344, Apr. 2018.
- [23] S. Wang, Y. Zhao, J. Xu, J. Yuan, and C.-H. Hsu, "Edge server placement in mobile edge computing," *J. Parallel Distrib. Comput.*, vol. 127, pp. 160–168, 2019.
- [24] X. He, R. Jin, and H. Dai, "Peace: Privacy-preserving and cost-efficient task offloading for mobile-edge computing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1814–1824, Mar. 2020.
- [25] Z. Zhou, H. Liao, X. Zhao, B. Ai, and M. Guizani, "Reliable task offloading for vehicular fog computing under information asymmetry and information uncertainty," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8322–8335, Sep. 2019.
- [26] F. Guo, H. Zhang, H. Ji, X. Li, and V. C. Leung, "An efficient computation offloading management scheme in the densely deployed small cell networks with mobile edge computing," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2651–2664, Dec. 2018.
- [27] P. A. Vikhar, "Evolutionary algorithms: A critical review and its future prospects," in *Proc. Int. Conf. Glob. Trends Signal Process., Inf. Comput. Commun.*, 2016, pp. 261–265.
- [28] H. Ye, G. Y. Li, and B.-H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.
- [29] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1243–1253, Feb. 2019.
- [30] M. Zhou, Y. Yu, and X. Qu, "Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: A reinforcement learning approach," *IEEE Trans. Intell. Transport Syst.*, vol. 21, no. 1, pp. 433–443, Jan. 2020.
- [31] N. Kumar, S. N. Swain, and C. Siva Ram Murthy, "A novel distributed Q-learning based resource reservation framework for facilitating D2D content access requests in LTE-A networks," *IEEE Trans. Netw. Serv. Manag.*, vol. 15, no. 2, pp. 718–731, Jun. 2018.

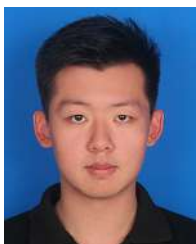


Xiaolong Xu received the Ph.D. degree in computer science and technology from Nanjing University, Nanjing, China, in 2016.

He was a Research Scholar with Michigan State University, East Lansing, MI, USA, from 2017 to 2018. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. He has authored or coauthored more than 80 peer-review articles in international journals and conferences. His research interests include edge computing, the Internet of Things (IoT), cloud computing, and big data.

Prof. Xu is a fellow of EAI (European Alliance for Innovation). He was the recipient of the Best Paper Award from the IEEE Cloud and Big Data (CBD) 2016, IEEE CPCSCom 2020, and International Conference on Security and Privacy in Digital Economy (SPDE) 2020.

Authorized licensed use limited to: Scms School Of Engineering And Technology. Downloaded on July 27, 2023 at 09:39:31 UTC from IEEE Xplore. Restrictions apply.



Bowen Shen is currently working toward the B.S. degree in computer science and technology with the School of Computer and Software at Nanjing University of Information Science and Technology, Nanjing, China.

His research interests include edge computing and IoT.



Sheng Ding received the graduate degree in computer science and technology, from East China University of Technology, Fuzhou, China, in 2003, and the master's degree in computer application technology from Ocean University of China, Qingdao, China, in 2012.

Since graduation, he has been working in the forefront of education, and has accumulated rich teaching experience. He has authored or coauthored many high-quality articles in various academic journals.



Gautam Srivastava (Senior Member, IEEE) received the Ph.D. degrees from the University of Victoria, Victoria, BC, Canada, in 2012.

In 2014, he joined a tenure-track position with Brandon University, Canada, and was promoted to the rank Associate Professor, in 2018. He has authored or coauthored more than 50 papers in high-impact conferences and high-status journals, including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (TWC), IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING

(TNSE), IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS (TCSS), IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS (TIA), and IEEE COMMUNICATIONS LETTERS. His research interests include social network and big data.

Prof. Srivastava was the recipient of the Best Oral Presenter Award in FSDM 2017. He is the Associate Editor for several international journals, such as IEEE TRANSACTIONS ON FUZZY SYSTEMS, IEEE ACCESS, etc. He has served as the leading Guest Editor FOR IEEE TRANSACTIONS ON FUZZY SYSTEMS.



Muhammad Bilal (Senior Member, IEEE) received the Ph.D. degree in information and communication network engineering from the School of Electronics and Telecommunications Research Institute (ETRI), Korea University of Science and Technology, Daejeon, South Korea, in 2017.

From 2017 to 2018, he was with Korea University, where he was a Postdoctoral Research Fellow with the Smart Quantum Communication Center. Since 2018, he has been an Assistant

Professor with the Division of Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies, Yongin, South Korea. His research interests include design and analysis of network protocols, network architecture, network security, the IoT, named data networking, blockchain, cryptology, and future Internet.

Dr. Bilal serves as an Editor for the IEEE FUTURE DIRECTIONS ETHICS AND POLICY IN TECHNOLOGY NEWSLETTER and the IEEE INTERNET POLICY NEWSLETTER.



Mohammad R. Khosravi received the B.Sc. degree from Shiraz Unuversity, Iran, in 2013, the M.Sc. degree fom Persian Gulf University, Iran, in 2015, and the Ph.D. degree from the Shiraz University of Technology, Iran, 2020, all in electrical engineering.

He is currently with the Department of Computer Engineering, Persian Gulf University, Bushehr, Iran, and has been with Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran. His main

interests include statistical signal and image processing, medical bioinformatics, radar imaging and satellite remote sensing, computer communications, industrial wireless sensor networks, underwater acoustic communications, information science, and scientometrics.



Varun G Menon (Senior Member, IEEE) He received the Ph.D. in computer science and engineering from Satyabhama University, India, in 2017.

He is currently an Associate Professor and Head of the Department of Computer Science and Engineering at the SCMS School of Engineering and Technology, Ernakulam, Kerala, India. His research interests include Internet of Things, 5G communications, fog computing and networking, underwater acoustic sensor net-

works, hijacked and predatory journals, ad hoc networks, opportunistic routing, and wireless sensor networks.

Dr. Menon was the recipient of the Top Peer Reviewer Award by Publons, in 2018 and 2019. He is a Distinguished Speaker of ACM. He is an Associate Editor for *Physical Communications* and *IET Quantum Communications*, Technical Editor for *Computer Communications*, and also an Editorial Board Member for IEEE FUTURE DIRECTIONS: TECHNOLOGY POLICY AND ETHICS. He has served over 20 conferences in leadership capacities including program Co-Chair, Track Chair, Session Chair, and Technical Program Committee member.



Mian Ahmad Jan received the Ph.D. degree in computer systems from the University of Technology Sydney (UTS), Sydney, Australia, in 2016.

He is currently an Assistant Professor with the department of computer science, Abdul Wali Khan University Mardan, Pakistan. His research interests include energy-efficient and secured communication in wireless sensor networks and internet of things, and has recently been actively involved in machine learning, big data analytics, smart cities infrastructure, and vehicular ad hoc networks.

Dr. Jan was the recipient of International Research Scholarship (IRS), UTS and Commonwealth Scientific Industrial Research Organization (CSIRO) scholarships. He was the recipient of the best researcher awarded for the year 2014 at the University of Technology Sydney Australia. He had been the recipient of various prestigious scholarships during his Ph.D. studies.



Mao-Li Wang received the B.S. degree in automation from Qufu Normal University, Jining, China, in 2004, and the M.S. and the Ph.D degrees in control theory and control engineering from Harbin Engineer University, China, in 2008.

He is a Professor with the School of Cyber Science and Engineering, Qufu Normal University. His research interests include Internet of Things, blockchain, and artificial intelligence.

[Home](#) > [Journal of Real-Time Image Processing](#) > [Article](#)

Special Issue Paper | [Published: 28 September 2020](#)

SD-Net: Understanding overcrowded scenes in real-time via an efficient dilated convolutional neural network

[Noman Khan](#), [Amin Ullah](#), [Ijaz Ul Haq](#), [Varun G. Menon](#) & [Sung Wook Baik](#) 

Journal of Real-Time Image Processing **18**, 1729–1743 (2021)

776 Accesses | **25** Citations | [Metrics](#)

Abstract

The advancements in computer vision-related technologies attract many researchers for surveillance applications, particularly involving the automated crowded scenes analysis such as crowd counting in a very congested scene. In crowd counting, the main goal is to count or estimate the number of people in a particular scene.

Understanding overcrowded scenes in real-time is important for instant responsive actions. However, it is a very difficult task due to some of the key challenges including clutter background, occlusion, variations in human pose and scale, and limited surveillance training data, that are inadequately

covered in the employed literature. To tackle these challenges, we introduce “SD-Net” an end-to-end CNN architecture, which produces real-time high quality density maps and effectively counts people in extremely overcrowded scenes. The proposed architecture consists of depthwise separable, standard, and dilated 2D convolutional layers. Depthwise separable and standard 2D convolutional layers are used to extract 2D features. Instead of using pooling layers, dilated 2D convolutional layers are employed that results in huge receptive fields and reduces the number of parameters. Our CNN architecture is evaluated using four publicly available crowd analysis datasets, demonstrating superiority over state-of-the-art in terms of accuracy and model size.

This is a preview of subscription content, [access via your institution.](#)

Access options

Buy article PDF

39,95 €

Price includes VAT (India)

Instant access to the full article PDF.

[Rent this article via DeepDyve.](#)

[Learn more about Institutional subscriptions](#)

References

1. Li, T., et al.: Crowded scene analysis: a survey. *IEEE Trans. Circuits Syst. Video Technol.* **25**(3), 367–386 (2014)
2. Hassaballah, M., Kenk, M.A., Elhenawy, I.M.: On-road vehicles detection using appearance and texture information. *Egypt. Comput. Sci. J.* **43**(1) (2019)
3. Zhang, C., et al.: Cross-scene crowd counting via deep convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015)
4. Zhang, C., et al.: Data-driven crowd understanding: a baseline for a large-scale crowd dataset. *IEEE Trans. Multimedia* **18**(6), 1048–1061 (2016)
5. Li, Y., Zhang, X., Chen, D.: Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In:

Proceedings of the IEEE conference on
computer vision and pattern recognition (2018)

6. Pan, J., et al.: Shallow and deep convolutional networks for saliency prediction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)

7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2015)

8. Wei, Y., et al.: Stc: A simple to complex framework for weakly-supervised semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(11), 2314–2320 (2016)

9. Wei, Y., et al.: Object region mining with adversarial erasing: a simple classification to semantic segmentation approach. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2017)

10. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint [arXiv:1511.07122](https://arxiv.org/abs/1511.07122) (2015)

11. Chen, L.-C., et al.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017)

12. Andri, R., et al.: YodaNN: An ultra-low power convolutional neural network accelerator based on binary weights. In: 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI). 2016. IEEE

13. Jia, Y., et al.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia (2014)

14. Qiu, J., et al.: Going deeper with embedded fpga platform for convolutional neural network. In: Proceedings of the 2016 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (2016)

15. Zhang, X., et al.: High-performance video content recognition with long-term recurrent

convolutional network for FPGA. In: 2017 27th International Conference on Field Programmable Logic and Applications (FPL). 2017. IEEE

16. Zhang, X., et al.: Machine learning on FPGAs to face the IoT revolution. In: 2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD). 2017. IEEE

17. Loy, C.C., et al.: Crowd counting and profiling: Methodology and evaluation. Modeling, simulation and visual analysis of crowds, pp. 347–382. Springer, Berlin (2013)

18. Dollar, P., et al.: Pedestrian detection: an evaluation of the state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(4), 743–761 (2011)

19. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). 2005. IEEE.

- 20.** Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comput. Vision* **57**(2), 137–154 (2004)
-
- 21.** Felzenszwalb, P.F., et al.: Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2009)
-
- 22.** Hassaballah, M., Awad, A.I.: Detection and description of image features: an introduction. *Image feature detectors and descriptors*, pp. 1–8. Springer, Berlin (2016)
-
- 23.** Chan, A.B., Vasconcelos, N.: Bayesian Poisson regression for crowd counting. In: 2009 IEEE 12th international conference on computer vision. 2009. IEEE.
-
- 24.** Idrees, H., et al.: Multi-source multi-scale counting in extremely dense crowd images. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2013)
-

25. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision. 1999. IEEE.

26. Lempitsky, V., Zisserman, A.: Learning to count objects in images. In: Advances in neural information processing systems (2010)

27. Pham, V.-Q., et al.: Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation. In: Proceedings of the IEEE International Conference on Computer Vision (2015)

28. Hassaballah, M., Awad, A.I.: Deep learning in computer vision: principles and applications. CRC Press, Boca Raton (2020)

29. Muhammad, K., et al.: Energy-efficient monitoring of fire scenes for intelligent networks. *IEEE Netw.* **34**(3), 108–115 (2020)

30. Ullah, A., et al.: Action recognition using optimized deep autoencoder and CNN for surveillance data streams of non-stationary environments. *Future Gener. Comput. Syst.* **96**, 386–397 (2019)

31. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems (2012)

32. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)

33. Yan, L., Zheng, Y., Cao, J.: Few-shot learning for short text classification. *Multimedia Tools Appl.* 77(22), 29799–29810 (2018)

34. Hassaballah, M., Hosny, K.M.: Recent advances in computer vision: theories and applications, vol. 804. Springer, Berlin (2018)

35. Ul Haq, I., et al.: Personalized movie summarization using deep cnn-assisted facial expression recognition. *Complexity*. **2019** (2019)

36. Muhammad, K., et al.: Deep learning for Multigrade Brain Tumor classification in smart healthcare systems: a prospective survey. *IEEE Trans. Neural Netw. Learn. Syst.* (2020)

37. Ullah, F.U.M., et al.: Violence detection using spatiotemporal features with 3D convolutional neural network. *Sensors* **19**(11), 2472 (2019)

38. Khan, S.U., et al.: Cover the violence: a novel deep-learning-based approach towards violence-detection in movies. *Appl. Sci.* **9**(22), 4963 (2019)

39. Walach, E., Wolf, L.: Learning to count with cnn boosting. In: *European conference on computer vision*. Springer, Berlin (2016)

40. Shang, C., Ai, H., Bai, B.: End-to-end crowd counting via joint learning local and global count. In: *2016 IEEE International Conference on Image Processing (ICIP)*. 2016. IEEE

41. Boominathan, L., Kruthiventi, S.S., Babu, R.V.: Crowdnet: A deep convolutional network for dense crowd counting. In: *Proceedings of the*

24th ACM international conference on
Multimedia (2016)

42. Marsden, M., et al.: Fully convolutional crowd counting on highly congested scenes. arXiv preprint [arXiv:1612.00220](https://arxiv.org/abs/1612.00220) (2016)

43. Sindagi, V.A., Patel, V.M.: Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2017. IEEE.

44. Zhang, Y., et al.: Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2016)

45. Onoro-Rubio, D., López-Sastre, R.J.: Towards perspective-free object counting with deep learning. In: European Conference on Computer Vision. Springer, Berlin (2016)

46. Shi, X., et al.: A real-time deep network for crowd counting. arXiv preprint [arXiv:2002.06515](https://arxiv.org/abs/2002.06515), (2020)

47. Wang, N., et al.: A light tracker for online multiple pedestrian tracking. *J. Real-Time Image Process.* 1–17
-
48. Balasundaram, A., Chellappan, C.: An intelligent video analytics model for abnormal event detection in online surveillance video. *J. Real-Time Image Process.* 1–16 (2018)
-
49. Shallari, I., Krug, S., O’Nils, M.: Communication and computation inter-effects in people counting using intelligence partitioning. *J. Real-Time Image Process.* 1–14 (2020)
-
50. Migniot, C., Ababsa, F.: Hybrid 3D–2D human tracking in a top view. *J. Real-Time Image Proc.* **11**(4), 769–784 (2016)
-
51. Poiesi, F., Cavallaro, A.: Predicting and recognizing human interactions in public spaces. *J. Real-Time Image Proc.* **10**(4), 785–803 (2015)
-
52. Nam, Y., Hong, S.: Real-time abnormal situation detection based on particle advection in crowded scenes. *J. Real-Time Image Proc.* **10**(4), 771–784 (2015)
-

53. Bahri, H., et al.: Real-time moving human detection using HOG and Fourier descriptor based on CUDA implementation. *J. Real-Time Image Process.* 1–16 (2019)
-
54. Chun, S., Lee, C.-S., Jang, J.-S.: Real-time smart lighting control using human motion tracking from depth camera. *J. Real-Time Image Proc.* **10**(4), 805–820 (2015)
-
55. Lotfi, M., Motamedi, S.A., Sharifian, S.: Time-based feedback-control framework for real-time video surveillance systems with utilization control. *J. Real-Time Image Proc.* **16**(4), 1301–1316 (2019)
-
56. Sam, D.B., Surya, S., Babu, R.V.: Switching convolutional neural network for crowd counting. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. IEEE
-
57. Sindagi, V.A., Patel, V.M.: Generating high-quality crowd density maps using contextual pyramid cnns. In: Proceedings of the IEEE International Conference on Computer Vision (2017)
-

58. Chan, A.B., Liang, Z.-S.J., Vasconcelos, N.: Privacy preserving crowd monitoring: Counting people without people models or tracking. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. 2008. IEEE
-
59. Sajjad, M., et al.: Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J. Comput. Sci.* **30**, 174–182 (2019)
-
60. Howard, A.G., et al.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017)
-
61. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition (2017)
-
62. Chen, L.-C., et al.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint [arXiv:1706.05587](https://arxiv.org/abs/1706.05587) (2017)
-
63. Zeiler, M.D., et al.: Deconvolutional networks. In: 2010 IEEE Computer Society Conference

on computer vision and pattern recognition.

2010. IEEE

64. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision (2015)

65. Lu, Z., et al.: The Classification of Gliomas Based on a Pyramid Dilated Convolution ResNet Model. *Pattern Recogn. Lett.* (2020)

66. Tota, K., Idrees, H.: Counting in dense crowds using deep features. In: *Proc. CRCV.* (2015)

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2019R1A2B5B01070067).

Author information

Authors and Affiliations

Intelligent Media Laboratory, Digital Contents Research Institute, Sejong University, Seoul, Republic of Korea

Noman Khan, Amin Ullah, Ijaz Ul Haq & Sung Wook Baik

**Department of Computer Science and
Engineering, SCMS School of Engineering
and Technology, Ernakulam, 683576, India**

Varun G. Menon

Corresponding author

Correspondence to [Sung Wook Baik](#).

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Khan, N., Ullah, A., Haq, I.U. *et al.* SD-Net: Understanding overcrowded scenes in real-time via an efficient dilated convolutional neural network. *J Real-Time Image Proc* **18**, 1729–1743 (2021). <https://doi.org/10.1007/s11554-020-01020-8>

Received	Accepted	Published
10 May 2020	13 September 2020	28 September 2020

Issue Date

October 2021

DOI

<https://doi.org/10.1007/s11554-020-01020-8>

Keywords

Crowd counting **Crowded scenes**

Deep learning

Dilated convolutional neural network

Real-time **Surveillance**

Received May 7, 2020, accepted July 29, 2020, date of publication August 6, 2020, date of current version August 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3014622

SafeCity: Toward Safe and Secured Data Management Design for IoT-Enabled Smart City Planning

HUI ZHANG^{1,2}, MUHAMMAD BABAR³, MUHAMMAD USMAN TARIQ³,
MIAN AHMAD JAN^{4,5}, VARUN G. MENON⁶, (Senior Member, IEEE),
AND XINGWANG LI⁷, (Senior Member, IEEE)

¹School of Energy Science and Engineering, Henan Polytechnic University, Jiaozuo 454003, China

²Coal Mining and Design Branch, China Coal Research Institute, Beijing 100013, China

³Department of Management, Abu Dhabi School of Management, Abu Dhabi, United Arab Emirates

⁴Informetrics Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁵Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁶Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India

⁷School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China

Corresponding author: Mian Ahmad Jan (mianjan@tdtu.edu.vn)

This work was supported in part by the National Natural Science Foundation General Fund under Project 51874109, in part by the Key Scientific and Technological Projects in Henan Province under Grant 182102310005, in part by the Science and Technology Support Plan of Guizhou Province (Science Support of Guizhou Province) under Grant [2019] 2861, in part by the Science and Technology Project for Outing and Young Talents of Guizhou (Talents of Science Platform in Guizhou) under Grant [2019] 5674, in part by the Key Scientific Research Projects of Higher Education Institutions in Henan Province under Grant 20A510007, and in part by the Fundamental Research Funds for the Universities of Henan Province under Grant NSFRF180309.

ABSTRACT The interaction among different Internet of Things (IoT) sensors and devices become massive and insecure over the Internet as we probe to smart cities. These heterogeneous devices produce an enormous amount of data that is vulnerable to various malicious threats. The generated data need to be processed and analyzed in a secure fashion to make smart decisions. The smart urban planning is becoming a reality through the mass information generated by the Internet of Things (IoT). This paper exhibits a novel architecture, SafeCity, that limelight the ecosystem of smart cities consists of cameras, sensors, and other real-world physical devices. SafeCity is a three-layer architecture, i.e., a data security layer, a data computational layer, and a decision-making layer. At the first layer, payload-based symmetric encryption is used to secure the data from intruders by exchanging only the authentic data among the physical devices. The second layer is used for the computation of secured data. Finally, the third layer extracts visions from data. The secured exchange of data is ensured by using Raspberry Pi boards while the computation of data is tested on trustworthy datasets, using the Hadoop platform. The assessments disclose that SafeCity presents precious insights into a secured smart city in the context of sensors based IoT environment.

INDEX TERMS Internet of Things, smart city, symmetric encryption, data management design, data analytics, data mining.

I. INTRODUCTION

Currently, 55% population of the world is in the cities that are expected to grow up to 67% by the year 2050 [1], [2]. The gradual increase in the urbanization poses various encounters for the decision-makers in proposing different facilities to the inhabitants of these cities. The ICT (Information and Communication Technologies) are used to make the cities smart enough by deploying and promoting sustainable devel-

The associate editor coordinating the review of this manuscript and approving it for publication was Jesús Hamilton Ortiz.

opment practices for addressing the growing challenges of urbanization. A solid foundation is offered for the Internet of Things (IoT) with an advancement in the field of smart cities' sensors by enabling them to interconnect [3]. Technology in the shape of smartphones, sensors, and other devices is playing a pivotal role in bringing the era of ubiquitous computing. In 2017, Gartner predicted that the number of interconnected devices will increase by 31% in 2017 by getting 8.5 billion and exceeded 20+ billion by the year 2020.

The IoT-enabled environment is a pattern where the processing of information is connected with every encountered

activity [4]. A huge number of real-world physical devices in a ubiquitous environment will generate voluminous data containing a variety of information that needs new forms of computation to facilitate enhanced decision making. The vast amount of data generated by the ubiquitous devices will add veracity, value, and variability to the Internet [5]. Advancement in the ubiquitous computing is causing in a large-scale valuable data or information, and with the assistance of Big Data tools and proficient machine learning methods, there is a great potential of analytical amenities to the smart cities [6]–[9]. A number of proposals are found to process and analyze the data generated by heterogeneous devices to perform efficient decision making.

Smart city data computation and pervasive intelligence expose the networks to security attacks, malware, and other cyber breaches. The inter-connectivity requirements of everyday physical devices would probably add numerous groundbreaking and resourceful malicious prototypes to IoT data computing [10]. The presence of malicious intruders may generate fabricated data to manipulate the sensed information of legitimate devices. The intruders may adversely affect the services and decision making in a ubiquitous environment. Furthermore, these malicious entities may liftoff attacks like denial-of-service by disrupting the transmission, and sensing of a ubiquitous environment to reduce the eminence of smart services [11].

Security provisioning in a ubiquitous environment is an intricate work since every machine possesses its identifiable unique characteristics and the uniqueness to be verified when connected to the Internet. The solutions for these ubiquitous devices in the marketplace lack the secured characteristics and are exposed to an extensive kind of adversarial attacks [12]. Besides, the existing privacy-preserving and authentication algorithms for smart ubiquitous environments involve complex and resource-intensive operations that require an abundance of resources. Most of these algorithms are not suitable for delay-sensitive and priority-based traffic generated in these environments.

In this article, we propose a safe and secured data management design for smart city planning using ubiquitous computing. The key contributions of the proposed architecture are as follows.

1. Payload-based symmetric encryption is proposed for a smart ubiquitous environment that is simple, lightweight, robust, and resilient against various malicious threats. The proposed approach uses 128-bit security primitives for secured exchange of data among the real-world physical devices.
2. A customized utility is proposed for the efficient loading of secured data into Hadoop. The proposed loading utility is efficient in terms of time and storage. The default HDFS (Hadoop Distributed File System) architecture is customized to achieve effective data storage. Our customized HDFS reduces storage consumption along with the network overhead.
3. The traditional YARN (Yet Another Resource Negotiator) Hadoop definition is customized for efficient data

computation. This is accomplished by introducing the concept of dynamic scheduling into the Hadoop YARN definition.

The remaining paper is ordered as follows. In Section 2, we spotlight the existing studies. In Section 3, we spotlight our proposed SafeCity framework for an IoT sensors based environment. In Section 4, the experimental results for secured data transmission and processing are presented. Finally, the paper is concluded in Section 5.

II. LITERATURE REVIEW

In this section, first we highlight the current works about the secure transmission of ubiquitous data collected from the smart cities, followed by their processing to extract valuable features.

A. SECURED TRANSMISSION OF DATA

Over the last decade, a lot of hype has been witnessed around building the concept of smart cities. Finally, the presence of sensor-embedded Internet of Things (IoT) platforms, ubiquitous connectivity, and cloud and data analytics has turned this concept into a reality. Although cities around the globe are seeking to become smarter, the applications of smart cities face a plethora of challenges in terms of security and privacy. These applications need to secure the gathered data from unauthorized access, disruption, annihilation, modification, inspection, and various other malevolent activities. In literature, numerous studies exist to protect the voluminous data traffic of smart ubiquitous cities from malicious entities. The error-prone communication channels used by the resource-starving sensors of smart cities limit the usage of TLS (Transport Layer Security) for seamless traffic flow [13]. As a result, most of the sensor nodes in smart ubiquitous environments rely on DTLS (Datagram Transport Layer Security) for the secured transmission of their data [14]. Nonetheless, the record layers of DTLS and handshake have a collective overhead of 25 bytes in each datagram header. The DTLS needs to be stripped of the resource-intensive operations to suit the resource-starving sensor nodes of smart cities [15].

In [16], the authors proposed an extremely lightweight encryption approach for the secured establishment of a unicast communication system in smart cities. The authors claimed that their model decreases the energy consumption and computational time of the sensor nodes. However, they did not provide any experimental and analytical results to verify their claim. In [17], the authors studied the use of DTLS for secured communication in a smart ubiquitous environment. They argued that the streaming applications of smart cities require an abundance of memory space and the use of DTLS is not feasible for them. The authors emphasized the use of compressed IPSec to offer security at the network layer for streaming applications.

A robust and resilient secured scheme for ubiquitous applications of smart cities was proposed in [18]. An RSA-based DTLS implementation was used for the secured exchange of ubiquitous data. However, both the RSA and DTLS have higher computational overheads due to resource-intensive

handshake mechanisms. The presence of complex cipher suites of RSA incurs a higher energy consumption and computational overhead for the ubiquitous operation of sensor nodes. The performance of the DTLS handshake was evaluated for ubiquitous smart devices using the Elliptic Curve Cryptography (ECC) [19].

In [20], a DTLS implementation for smartphones was proposed using the Constrained Application Protocol (CoAP). The proposed scheme involves computationally difficult encryption suites, requires ample processing and power memory, and is not suitable for sensor nodes of the smart cities. In [21], a lightweight encryption approach was proposed for ubiquitous communication in a smart city environment. Prior to establishing a secured session, the proposed approach validates the identities of clients and servers. For authentication, symmetric encryption with 128-bit security primitives were used. However, the proposed scheme is not validated experimentally to verify its efficiency, robustness, and resilience.

B. DATA PROCESSING AND FEATURE EXTRACTION

In this section, the challenges and issues in the existing works for smart city planning utilizing the Big Data analytical techniques are presented. In [22], the authors designed a model to compute Big Data generated in the IoT-based smart health setting. It involves the separation of vigorous data into subclasses that are based on hypothetical simulation of data fusion to improve computational effectiveness. The key issues underlined in this model are the use of customary MapReduce Cluster management for Apache Hadoop server, insufficient data loading to Hadoop, a conceptual framework, and the utilization of only healthcare datasets.

A Big Data analytics framework comprised of various tiers was proposed for urban planning in [23]. Each tier of the framework is responsible for different activities of the Big Data analytics to have efficient modularization of the overall process. Although, it is a complete framework from data generation and collection to application and usage of the analyzed data, it causes significant delay in processing and the use of classical MapReduce deteriorates the performance [24]. Moreover, prior to data loading, the authors focused on data aggregation while overlooking the data loading competence.

An IoT-enabled framework using Hadoop-based Big Data analytics was proposed in [25] for a smart city application. The proposed framework has different layers from data acquisition to the application. The main problem of this framework is that the data loading efficiency was ignored.

A proposal based on the analysis of Big Data that endorses the perception of SCC (smart and connected societies) for smart cities was proposed in [26]. The SCC model is a conceptual framework that was not implemented. A similar model was proposed for the ubiquitous smart city application in [27]. However, this model was not implemented as well. Moreover, [26] and [27] overlooked the data loading and ingestion into a distributed ubiquitous smart city environment. In addition, many solutions have been proposed to treat similar problems of Big Data analytics in smart ubiquitous 145258

environments [28], [29]. Vecular fog computing may also be utilized for smart city planning [30]. However, a critical issue in the design of these methods is the deployment of a traditional cluster resource management scheme and insufficient data loading to the Hadoop server.

A graph-oriented architecture to analyze the Big Data in a smart ubiquitous transportation system was proposed in [31]. This graph-based solution is more scalable and efficient, but it incurs additional delay due to graph processing. In addition to processing delay, the proposed solution was tested only for the transportation dataset, and loading the Big Data to the Hadoop server and its efficiency was overlooked. The proposed architecture was tested only for a healthcare dataset. The authors proposed a multi-level data processing scheme, based on parallel processing, for Big Data analysis. However, a YARN-enabled solution was provided but the data ingestion efficacy was ignored.

III. A SAFE AND SECURED DATA MANAGEMENT FRAMEWORK

For a smart and safe city to perform intelligent and secure decisions, the ubiquitous data collected by the devices are processed using different approaches. In SafeCity, the data analysis and machine learning approaches are applied to the data generated and acquired in a ubiquitous environment. The acquisition is carried out by systems that convert the analog information into digital. The cellular technology, i.e. 4G/LTE, is used as a bridging technology between the users, devices, and the system, as shown in Figure 1.

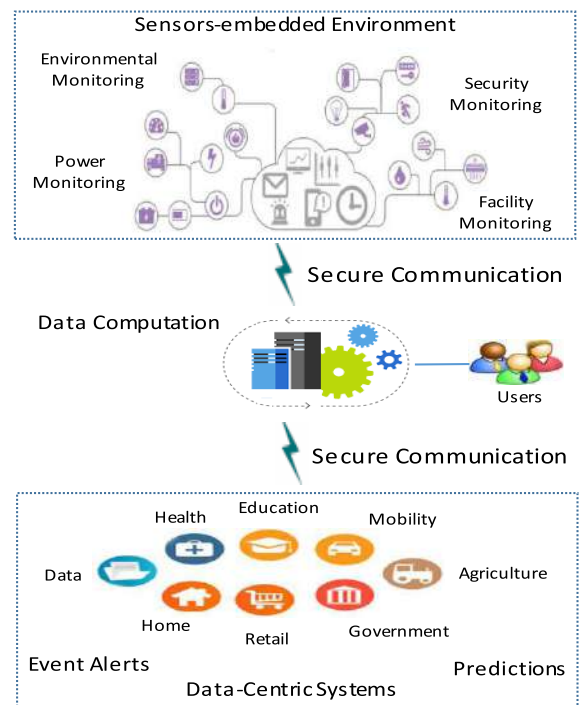


FIGURE 1. Overview of the proposed system.

To design a ubiquitous environment, numerous surveillance cameras, wired and wireless sensors, and device-

mounted sensors are deployed. Data sensing, acquisition, and collection are performed in this environment. Digital loggers and digital data acquisition systems are used to detect and collect data from devices and disseminate them with the help of the Internet. The produced ubiquitous data are secured before forwarding to a computational unit for safe and secured processing and transmission. Afterward, the decisions are made on the secured ubiquitous data. The proposed system is a three-layer architecture, i.e., a ubiquitous data security layer, a ubiquitous data computation layer, and a decision-making layer. A payload-based authentication approach is utilized in the first layer to make the ubiquitous data secured from adversaries.

This layer ensures that only secured data is forwarded. The second layer is accountable for the resource-intensive processing of secured ubiquitous data at the conventional computing platforms. Finally, the third layer provides insights from the ubiquitous data and makes smart decisions. The proposed architecture is shown in Figure 2. The comprehensive description of each layer is given in the following subsections.

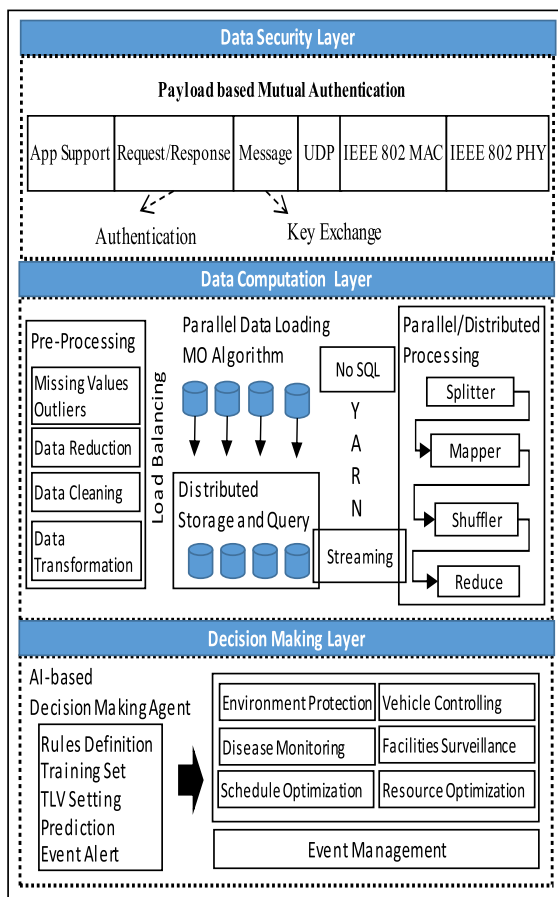


FIGURE 2. System architecture of SafeCity.

A. DATA SECURITY LAYER

This layer of SafeCity is linked to the data sources. The data received from the sensors are in the form of messages. At this layer, message identification and authentication

are performed using a simple payload-based authentication scheme. The proposed scheme uses the CoAP protocol [32] for message exchange and authentication at the application layer of each data source. In ubiquitous environments, most of the CoAP-based solutions are relied on the use of DTLS to ensure the protected transfer of resources between the devices. However, the DTLS-enabled CoAP stack incurs an excessive computational and communication overhead. Furthermore, the use of DTLS in combination with CoAP adds an extra layer of protocol header for security provisioning. In our approach, the security of data messages is not compromised while transferred between clients and servers. The session key is transmitted within the payload messages while authentication is achieved at the request-response communication, as shown by the top layer in Figure 2. In SafeCity, CoAP is equipped with secured features for authentication, efficiency, robustness, and defense against a number of malevolent threats.

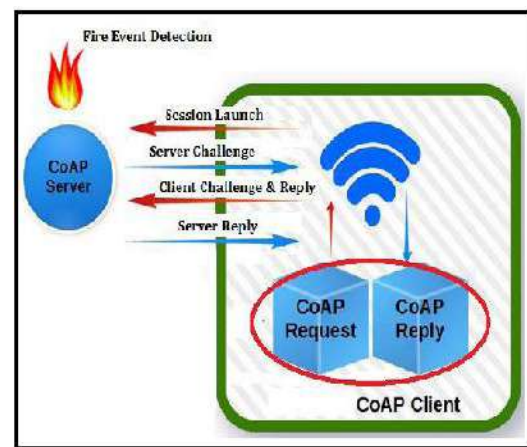


FIGURE 3. Mutual authentication.

During the authentication process, the resource-constrained clients communicate with a server to verify each other identities. As an example, the ubiquitous clients of Figure 1 observe various events such as, temperature, humidity, pollution, and fire eruption, at the server. For a server to provide access to the residing resources, both the parties need to be mutually authenticated. In SafeCity, the authentication is accomplished using four handshake messages. A maximum of 256-bits is used within the payload of each message. The four handshake messages are session launch, server challenge, client challenge and reply, and server reply, as shown in Figure 3. The session launch is headed by a provisioning stage where the clients share a secret key with the server. The server conserves a trace of keys, based on an associated unique identifier (ID). The exchange of a session key between the client and server takes place upon successful authentication. For each client, a session key is implanted on the device at the manufacturing time. If an impostor strives to raze the client, a specific alarm is spawned to notify the crack. To encode the payload of authentication information, the Advanced Encryption Standard (AES) is utilized.

During the session launch, a secret key λ_i is shared with the server, where λ_i is 128-bit long. The λ_i is identified only by client_{*i*} (it belongs) and server, where $i \in \{1,2,3,\dots,I\}$. Each i has a unique identifier that helps the server to execute a look-up table for verification of identity. The session launch is similar to a Hello message and its payload consists of CoAP options fields, i.e., **Auth** and **Auth-Msg-Type**, to indicate the type of operations performed between the client and a server. After the session launch, the next step is the server challenge, in which the server creates a challenge for the client. The encounter containing a pseudo-random nonce η_r and a session key μ produced by the server. The following equations are used to create a challenge.

$$\vartheta = \lambda_i \oplus \mu \tag{1}$$

$$C_r = \text{AES}\{\lambda_i, (\vartheta|\eta_r)\} \tag{2}$$

where, i is the ID of a client, ϑ is the intermediate value generated by the server, and C_r is the challenge sent to i . In the client challenge and reply message, the client retrieves η_r and λ from the server challenge and creates a challenge in response using the following equations.

$$\vartheta' = \eta_r \oplus \lambda_i \tag{3}$$

$$C_i = \text{AES}\{\mu, (\vartheta'|\eta_i)\} \tag{4}$$

where η_i is the pseudo-random nonce and ϑ' is the intermediate value generated by the client, and C_i is the challenge sent to the server. Upon receiving, the server tries to retrieve η_r from the client's challenge. If this nonce is present, the status of i changes to **Authenticated**, and the server responds to the client's challenge to complete the authentication process, using the following equation.

$$C_r = \text{AES}\{\lambda_i, (\eta_r|\mu)\} \tag{5}$$

B. DATA PROCESSING AND COMPUTATION LAYER

Versatile analysis and intelligent processing on huge data streams can be unrealistic and infeasible if the data streams are not properly pre-processed. Data pre-processing are performed prior to the core computation and processing. The pre-processing steps involve the reduction to realize the reduced data with similar properties, data transformation to standardize data to an appropriate arrangement for processing, and data cleansing. These activities are carried out using machine learning approaches. The objective is to dig out the data about various sets of an IoT domain, based on its characteristics. Next, the data loading is performed using multiple attribute criteria model (MACM) in the context of the Hadoop ecosystem. The MACM includes parallel data loading using the customized utility. The HDFS saves the huge files in small chunks that are customized to avoid too much data and metadata, that would otherwise create the overhead.

In HDFS, a replication method is to replicate the original chunk of data which is a time-consuming task. As a result, customized replication is proposed in this paper. Moreover, the Sqoop utility is used due to the parallel loading of data

using the map method. The proposed scheme utilizes Sqoop that offers connectivity to the external databases. The utilization of Sqoop brings a variety of features in SafeCity, such as loading with increments, complete import, parallel import, and corresponding export, compression, easy movement, enterprise independence, and auto-generation of tedious user side's code. Data processing and analytics are carried out using the MapReduce programming paradigm. Hadoop divides input dataset into small blocks of same size files, known as input splits. The size of the split is usually identical to the block or chunk size. One specific task (known as map task) is formed for each split that performs the function of the map, defined by the programmer, for each row (a record). A RecordReader is used to arrange the rows as a pair (key-value). The MapReduce process is depicted in Figure 4. The outputs of the map are not stored in HDFS, these results are stored in the local storage. Results from a number of mappers are the input for the reduce task. Reduce tasks do not include the advantage of data locality characteristic. Therefore, the stored map results have to transfer crossway the system to that specific location, where the job of reducing is performing. The of the reducer result is stored on HDFS.

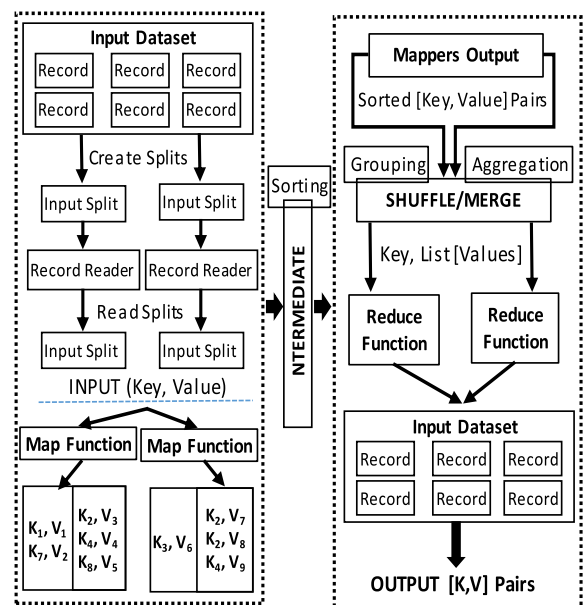


FIGURE 4. MapReduce paradigm.

Our projected scheme is grounded on the up-to-date depiction of Apache Hadoop framework which is embedded with Yet Another Resource Negotiator (YARN) and is accountable for data computation and cluster management. Unlike conventional MapReduce, the computation elements and resource management is separated by YARN. The YARN-enabled model is not limited to the MapReduce classical mechanism. The YARN is preferred due to limitations of classical MapReduce that are mostly associated with scalability and workload support. In the proposed architecture,

YARN has a ResourceManager that runs as a master daemon by managing the accessible cluster resources among a wide range of competing and contending applications. The ResourceManager keeps track of the available resources and live nodes on the cluster. As it is the solo process having this information, so it coordinates the resource allocation and scheduling between the submitted applications. The allocation decisions are made in a secured, multi-tenant, and shared way, e.g. based on queuing capacity, data locality, an application priority, etc. On the submission of an application, a lightweight process instance, also known as ApplicationMaster, is initiated that is responsible for the execution of all the tasks within an application. It is comprised of tasks monitoring, restarting failed tasks, and calculating the overall values of the used application counters. In the existing literature, the classical MapReduce framework is utilized where a single JobTracker is responsible to take care of these responsibilities for all the jobs. Utilizing a single JobTracker in huge clusters exposes them to the scalability bottleneck.

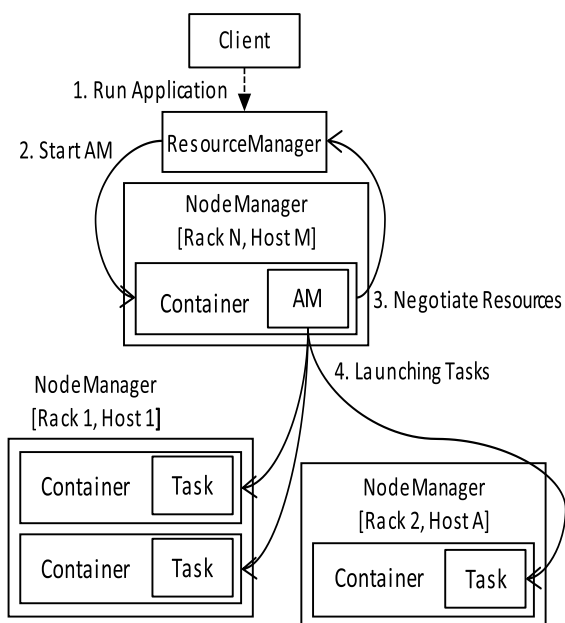


FIGURE 5. Yet another resource negotiator (YARN).

Different tasks associated with a particular application and an ApplicationMaster are controlled, monitored, and managed by the corresponding NodeManagers. Unlike the TaskTracker of a classical MapReduce framework, NodeManager is an efficient and more generic version of the TaskTracker. The NodeManager has many resource containers that are created dynamically, rather than having a defined number of slots (maps and reduces). All the components of the YARN such as ResourceManager, NodeManagers, ApplicationMaster, and containers cooperate with each other in a specific way upon the submission of an application in the cluster of YARN. This interaction of different parts of a YARN framework is shown in Figure 5.

The application is submitted using the Hadoop jar command in CLI or using Java IDE to RM, in a similar way to classical MR. A complete list of running jobs on the Hadoop cluster and all the available and accessible resources on every NM (live) are maintained by RM. The RM needs to decide which application is the next to acquire a piece of cluster resource. A number of constraints are taken into consideration while taking this decision such as fairness and capacity of the queue. The RM employs a scheduler that focuses mainly on scheduling activities. It deals with accessing the resources of a cluster and decides when and who will access them. Within an application, the task monitoring is not carried out by the scheduler and it never tries to restart a failed task. When the submission of a new application is accepted by ResourceManager, first the scheduler decides to select a container where ApplicationMaster will be started and run.

The ApplicationMaster will be in charge of the entire life cycle of the application when it starts. Primarily, ApplicationMaster would be requesting for various resources to the overall manager (ResourceManager) in order to inquire for different containers that are required to execute tasks of a particular application. A request for a particular resource is just a demand for several containers to assure various resource necessities, i.e., a number of resources. For example, CPU share, MB memory, preferred location, e.g. rack name, hostname or if no preference is required then * is used, and priority inside the current application.

The ResourceManager grants a container, whenever possible, that satisfies the request made by an ApplicationMaster. On a specific host, the application is permitted by the container to utilize specified resources. ApplicationMaster requests the NodeManager to launch an application-specific task to utilize these resources after a container is granted.

Please recall that the NodeManager is responsible to manage the host on which a particular container is assigned. The application-specific task could be any particular task written in any framework, e.g. MapReduce. The NodeManager only monitors and examines the resource usage in the containers. It does not monitor the tasks and destroys them if they use more than the allocated memory.

The ApplicationMaster is responsible for monitoring the restarting tasks in fresh containers that are failed, the progress of tasks and its application, and provides the progress back to a client. The ApplicationMaster closes itself and releases its container on completion of the application. Nevertheless, the RM does not check the tasks inside an application at all. It only confirms the health of the ApplicationMasters. In this paper, a flowchart is proposed using the MapReduce programming paradigm that is applied to a water dataset. This flowchart is used to collect the values/quantity of water consumption against different houses to govern the level of water and its demand. The pictorial illustration of recommended MapReduce is depicted in Figure 6.

The mapper gets the offset of a line as a specific key and the entire row is considered as a value. The time parameter (timestamp) and associate values are produced as output

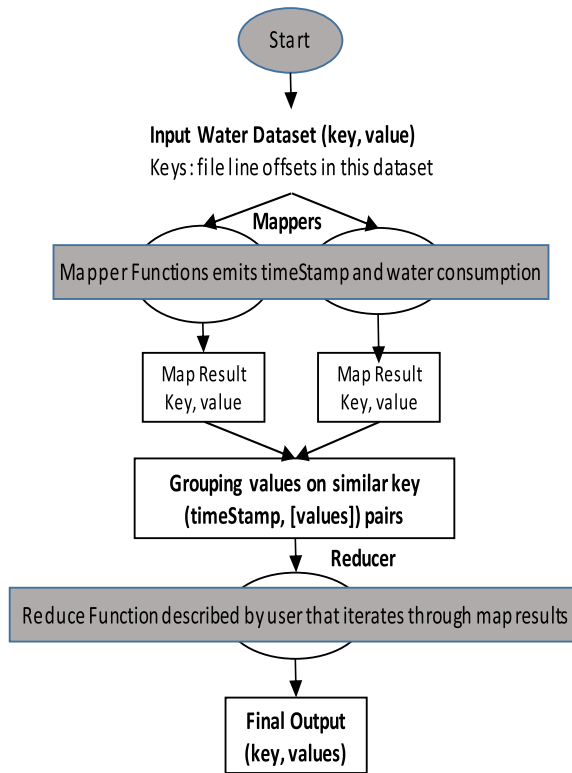


FIGURE 6. MapReduce flowchart for water dataset.

Algorithm 1 Mapper for Water Dataset

```

BEGIN
Input:
  key: line-offset
  value: = row
Output:
  key: facilityID
  value: LOTLINK
  //containing water consumption measurement

  // line splitting
  facilityID, LOTLINK: = line.split('\t')
  key: = facilityID
  value: = LOTLINK
  emit (key, value)
END
  
```

by the mapper. The reducer clusters the necessary associate values alongside every timeStamps and relates with the TLV (threshold limit value). Information with regard to the water consumption of different houses is obtained with the help of such algorithms. As the MapReduce executes various jobs in 2 phases, i.e., Map phase and Reduce phase, therefore, a separate Map function and a Reduce function is proposed for the flowchart of Figure 6. In Algorithm 1, we present the mapper for the water dataset and in Algorithm 2, we present the reducer for the same dataset.

Algorithm 2 Reducer for Water Dataset

```

BEGIN
Input:
  key: facilityID
  value: LOTLINK
Output:
  key: facilityID
  value: LOTLINK greater than threshold
initialize threshold
final []
FOR each (LOTLINK) at facilityID DO
IF (LOTLINK > threshold)
Begin
  final.append (LOTLINK)
  key: = facilityID
  value: = final
  emit (key, vaue)
End IF
END
  
```

C. DECISION-MKING LAYER

The intelligent decision making is the key to our SafeCity framework that includes the prediction, creation of training sets, thresholds setting, rules definition, and event management. It acts as the moderator between the end-users and it is carried out by the decision-making agent, based on AI approaches. Various limits are defined and several rules are set for the assessment of different datasets. The processing of data is carried out using these rules according to proposed algorithms. The TLV (Threshold Limit Value) is a precise value set for each dataset also known as threshold or limit which is the base for event generation and decision making. Likewise, several rules are set centered on corresponding limits in the form of if/then statements that are utilized for decision making. The notification and event alert component determines the specific recipient of a generated event. Hence, it notifies the operator with the generated event for further actions.

IV. SYSTEM EVALUATION AND ANALYSIS

The detailed analysis and discussion of results achieved using SafeCity discuss in this segment. The secured data authentication is realized using Raspberry Pi boards for the client-server interface model. The Libcoap library is used for Raspbian operation system that provides basic communication among the ubiquitous devices. The analysis is carried out on a dataset that is realistic to evaluate the SafeCity scheme using the premeditated algorithms. The implementation of our ubiquitous data computation layer is carried out using the Hadoop cluster on Ubuntu OS along with Sqoop. Moreover, Java is used for the MapReduce implementation by utilizing the pre-defined classes (mapper and reducer). The data is received from diverse but trustworthy sources that

are authentic. These datasets contain the transportation data, i.e., vehicles on roads in Aarhus city, Denmark. The water dataset homes are gained from the houses in Surrey, Canada.

A. SYSTEM EVALUATION FOR SAFETY AND SECURITY

The experimental results concerning the ubiquitous data security layer are illustrated here. A comparison of our payload-based authentication for SafeCity and CoAP-based DTLS implementation for smartphones is provided in Figure 7. DTLS+ denotes a smartphone (ubiquitous device) operating as a server and a workstation as a client. On the other hand, DTLS* denotes the handshake between a smartphone and a workstation, where the smartphone operates as a client and the workstation as a server. As the figure shows, SafeCity has a much lower handshake duration and standard deviation in comparison to DTLS* and DTLS+.

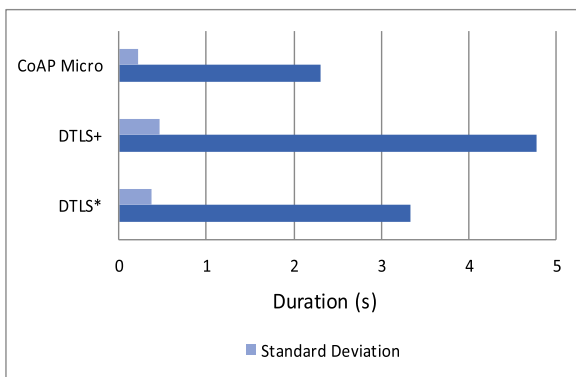


FIGURE 7. Handshake duration.

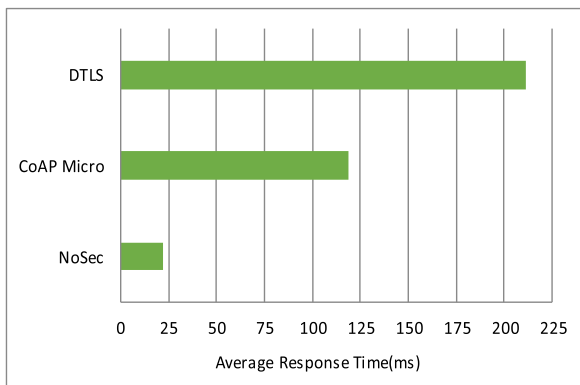


FIGURE 8. Average response time.

Similarly, SafeCity focuses on asynchronous communication of CoAP messages over the UDP sockets. A record of transferred Confirmable (CON) requests is maintained by every client. The mean reaction time for one CON request message of 1 byte is compared with DTLS exchange and the CoAP protocol with no added security, in Figure 8. SafeCity has a much lower average response time in comparison to DTLS because the latter involves computationally complex cipher suites and a resource-intensive record layer. CoAP with no added security has a slower response time but it is prone to various malicious and adversarial attacks.

TABLE 1. Average consumption (kb).

CoAP Micro	HTTP	HTTP/U DP	CoAPBlip	TinyCoAP
207	802	4009	7160	8498

The memory utilization of a CON request is evaluated at the compile time in Table 1. The proposed SafeCity is compared with the existing schemes for a CON message of minimum 500 bytes, as depicted in Figure 9 too. Among the current schemes, CoAPBlip [33] allocates considerable storage to messages at the compile time of the message. TinyCoAP [34] is a variation of the standard libraries of C that need the TinyOS element for its installation on a ubiquitous device. HTTP has a short foot-print of memory as it doesn't offer a trustworthiness method or correlation of a request/response. Both TinyCoAP and CoAPBlip use resource-consuming libraries and have a much higher memory consumption.

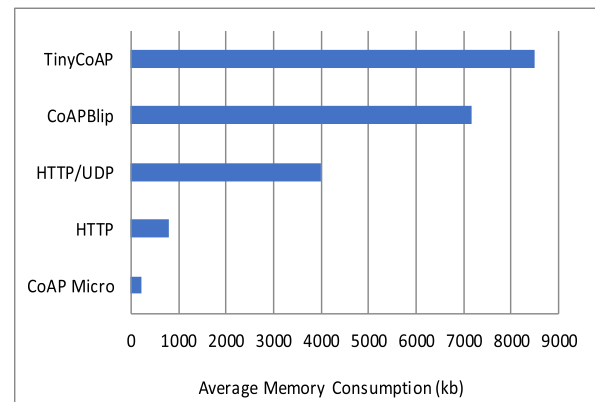


FIGURE 9. Average memory consumption.

B. SYSTEM EVALUATION FOR DATA PROCESSING AND COMPUTATION

Our SafeCity architecture generates alerts in real-time for a particular ubiquitous environment. In this section, we evaluate SafeCity in terms of efficiency by considering the execution time and throughput. To examine the system performance in real-time, various datasets, such as vehicular and water, are replayed to our Hadoop-based YARN framework of SafeCity. The throughput is assessed using datasets by increasing the data size. The efficiency concerning throughput is measured as shown in Figure 10. It can be observed that with the growth in size, the processing speed is reduced. The system throughput of Yarn-based framework is considerably higher in comparison to the existing classical MR-based solution.

Table 2 reveals the processing time, also known as the execution time proposed framework in the context of data volume. The execution time is evaluated for different sizes

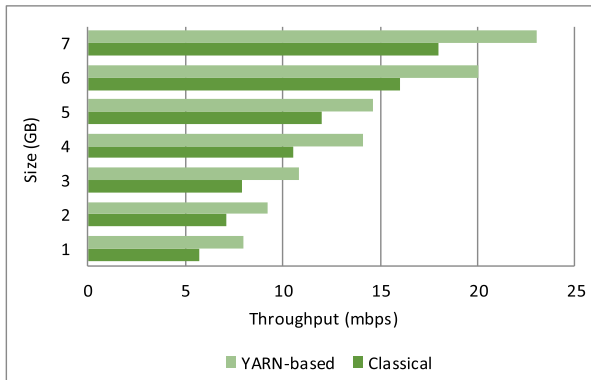


FIGURE 10. System throughput.

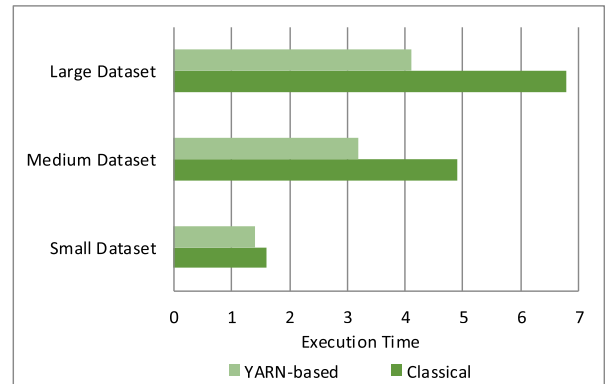


FIGURE 11. Execution time (s).

TABLE 2. Processing time of proposed framework.

Size (GB)	Time (ms)
1	67
2	78
3	96
4	119
5	133.5
6	150.9
7	168.3
8	185.7
9	203.1
10	220.5
11	237.9
12	255.3
13	270

TABLE 3. Average consumption (kb).

	Minor	Average	Huge
Traditional	1.4	4.9	6.8
Proposed	1.6	3.2	4.1

of data. The data size is started from 500MB and experienced up to 13 GB of data.

Table 3 determines the processing time in comparison to the classical structure. The time is calculated for minor, average, and huge datasets. It is observed that the processing time improves when the dataset size is increased. Figure 11 demonstrates the execution time of jobs using our Yarn-based framework in comparison to the existing scheme. The execution time is evaluated for small, medium, and large datasets. It is observed that the processing time improves when the dataset size is increased. It is mostly because of the data loading efficiency and improvement.

C. DATA ANALYSIS

The time difference of data loading is not perceptible when the size is smaller. The data ingestion time is pretty evident when the bulk of a dataset is larger due to the

replication approach. The query that arises is the threshold data, to discover the TLV size, the data loading performance is measured by testing the different sizes of data.

The TLV size is the point where the time difference becomes positive (greater than 0) which means a significant change occurs. The TLVs for various attributes are set using the outputs of similar trials. Taking into account the data ingestion tool experiments, the TLV size is 900MB (size of data). At this value, the effect of the data ingestion period is experienced as shown in Figure 12. This figure demonstrates that 1GB of size does not generate any change even if the automated ingestion is practice. The productivity is attained when dataset size is greater than 900 MB at least.

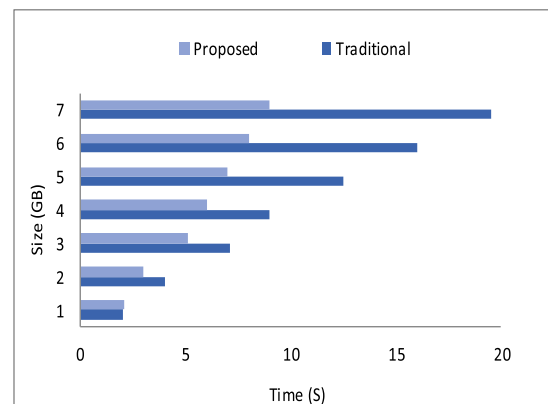


FIGURE 12. Data loading efficiency.

The water consumption is evaluated to achieve sustainable water management in the city due to the inconsistent consumption of water could be a disaster in the future. The data utilized in our research contains information about the city of Surrey, Canada. It comprises of the water intake of the houses in Surrey that is processed using our proposed algorithms. The results are demonstrated in Figure 13.

It shows the houses consumed more than 82000 liters each month. The defined TLV is 82000 found from the rule engine. The water usage higher than the TLV is particularly highlighted in this figure and this can cause frightening

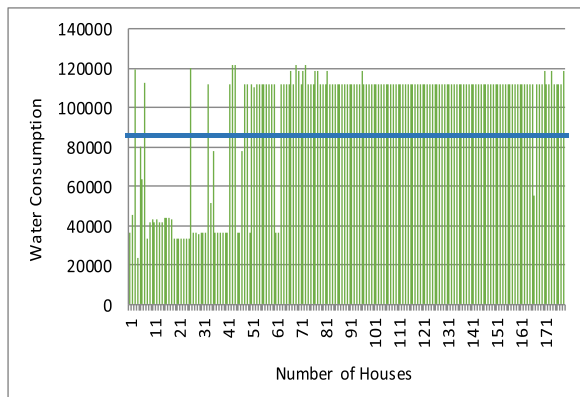


FIGURE 13. Water consumption.

situations for the authorities. It is observed that almost 50% of the consumers consumed more than the threshold limit. Most of the consumers, above the TLV limit, consumed water between 110000 to 120000 liter, which is quite alarming. Up-to-date fabrication methods could be industrialized to control the issues of the consumers in a city.

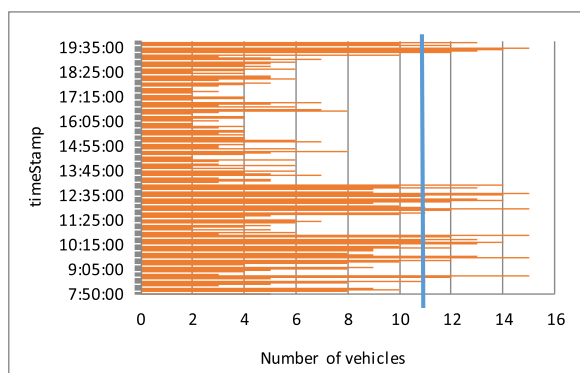


FIGURE 14. Number of vehicles on the road.

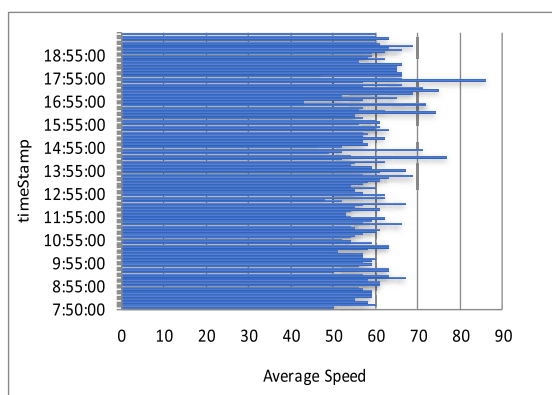


FIGURE 15. Average speed of automobiles.

Regarding traffic management, we consider the traffic data about road congestion. The data is intelligently processed using the SafeCity framework to overcome the traffic issues when the vehicles on roads surpass TLV. Figure 14 reveals

the vehicles and the corresponding TLV. It depicts vehicles at a different time on the roads. It is observed that due to schooling hours, there are more cars between 8:05-12:15 PM due to school and office timing in the city.

Furthermore, the average speed of vehicles is revealed in Figure 15. It is noticed that the average speed of the vehicles is quite alike all day, except from 13:00 to 18:00, when there are few vehicles.

V. CONCLUSION

This paper has envisioned the vital role of safety and security in IoT-enabled data computation and communication to achieve safe and secure decisions. The data generated by IoT sensors exploit the association between various features of data and enables the meaning of a safe city. We have suggested the conception of SafeCity and proven its applicability using apache and Hadoop, via cautious investigation and assessment of the presence of residents in the evolving smart cities. SafeCity carefully controls the encounter of security and computation faced by the ubiquitous data. It is a layered architecture that is composed of a data security layer, data computation layer, and decision-making layer. A payload-based authentication approach is utilized at the ubiquitous data security layer to secure the ubiquitous data from malevolent entities.

The data computation layer is liable for the processing of secured data. Finally, the decision-making layer extracts insights for making smart decisions. The ubiquitous data security is evaluated using the Raspberry Pi boards while the ubiquitous data computation is tested on trustworthy datasets, using Hadoop. In association with the current methods, SafeCity is trivial about handshake duration, response time, and average memory consumption. Furthermore, it attains a lesser processing time, greater throughput, and efficient about massive data ingestion.

REFERENCES

- [1] P. Bocquier, "World urbanization prospects: An alternative to the UN model of projection compatible with the mobility transition theory," *Demograph. Res.*, vol. 12, pp. 197–236, May 2005.
- [2] J. L. Hernández, R. García, J. Schonowski, D. Atlan, G. Chanson, and T. Ruohomäki, "Interoperable open specifications framework for the implementation of standardized urban platforms," *Sensors*, vol. 20, no. 8, p. 2402, Apr. 2020.
- [3] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.
- [4] M. Weber and I. P. Žarko, "A regulatory view on smart city services," *Sensors*, vol. 19, no. 2, p. 415, 2019.
- [5] A. Entezami, H. Sarmadi, B. Behkamal, and S. Mariani, "Big data analytics and structural health monitoring: A statistical pattern recognition-based approach," *Sensors*, vol. 20, no. 8, p. 2328, Apr. 2020.
- [6] M. Babar and F. Arif, "Smart urban planning using big data analytics based Internet of Things," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput., ACM Int. Symp. Wearable Comput.*, Sep. 2017, pp. 397–402.
- [7] M. Babar and F. Arif, "Smart urban planning using big data analytics to contend with the interoperability in Internet of Things," *Future Gener. Comput. Syst.*, vol. 77, pp. 65–76, Dec. 2017.
- [8] M. Babar and F. Arif, "Real-time data processing scheme using big data analytics in Internet of Things based smart transportation environment," *J. Ambient Intell. Hum. Comput.*, vol. 10, no. 10, pp. 4167–4177, Oct. 2019.

- [9] M. Babar, A. Rahman, F. Arif, and G. Jeon, "Energy-harvesting based on Internet of Things and big data analytics for smart health monitoring," *Sustain. Comput., Informat. Syst.*, vol. 20, pp. 155–164, Dec. 2018.
- [10] J. Granjal, E. Monteiro, and J. S. Silva, "Security for the Internet of Things: A survey of existing protocols and open research issues," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1294–1312, 3rd Quart., 2015.
- [11] S. Rajesh, V. Paul, V. Menon, and M. Khosravi, "A secure and efficient lightweight symmetric encryption scheme for transfer of text files between embedded IoT devices," *Symmetry*, vol. 11, no. 2, p. 293, Feb. 2019.
- [12] M. A. Jan, F. Khan, M. Alam, and M. Usman, "A payload-based mutual authentication scheme for Internet of Things," *Future Gener. Comput. Syst.*, vol. 92, pp. 1028–1039, Mar. 2019.
- [13] A. Vėnčauskas, N. Morkevicius, V. Jukavičius, R. Damaševičius, J. Toldinas, and Š. Grigaliūnas, "An edge-fog secure self-authenticable data transfer protocol," *Sensors*, vol. 19, no. 16, p. 3612, Aug. 2019.
- [14] S. L. Keoh, S. S. Kumar, and H. Tschofenig, "Securing the Internet of Things: A standardization perspective," *IEEE Internet Things J.*, vol. 1, no. 3, pp. 265–275, Jun. 2014.
- [15] S. Raza, L. Seitz, D. Sitenkov, and G. Selander, "S3K: Scalable security with symmetric keys—DTLS key establishment for the Internet of Things," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 3, pp. 1270–1280, Jul. 2016.
- [16] A. Bhattacharyya, A. Ukil, T. Bose, and A. Pal. *Lightweight Mutual Authentication for CoAP (WIP)*. Accessed: Mar. 3, 2014. [Online]. Available: <https://draft-bhattacharyya-core-coap-lite-auth-00>
- [17] J. Granjal, E. Monteiro, and J. S. Silva, "On the feasibility of secure application-layer communications on the Web of things," in *Proc. 37th Annu. IEEE Conf. Local Comput. Netw.*, Oct. 2012, pp. 228–231.
- [18] T. Kothmayr, C. Schmitt, W. Hu, M. Brünig, and G. Carle, "DTLS based security and two-way authentication for the Internet of Things," *Ad Hoc Netw.*, vol. 11, no. 8, pp. 2710–2723, Nov. 2013.
- [19] J. Granjal, E. Monteiro, and J. S. Silva, "On the effectiveness of end-to-end security for Internet-integrated sensing applications," in *Proc. IEEE Int. Conf. Green Comput. Commun. (Green-Com)*, Nov. 2012, pp. 87–93.
- [20] D. Tralbalza, S. Raza, and T. Voigt, "Indigo: Secure coap for smartphones," in *Wireless Sensor Networks for Developing Countries*. Berlin, Germany: Springer, 2013, pp. 108–119.
- [21] M. A. Jan, P. Nanda, X. He, Z. Tan, and R. P. Liu, "A robust authentication scheme for observing resources in the Internet of Things environment," in *Proc. IEEE 13th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Sep. 2014, pp. 205–211.
- [22] S. Din, H. Ghayvat, A. Paul, A. Ahmad, M. M. Rathore, and I. Shafi, "An architecture to analyze big data in the Internet of Things," in *Proc. 9th Int. Conf. Sens. Technol. (ICST)*, Dec. 2015, pp. 677–682.
- [23] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using big data analytics," *Comput. Netw.*, vol. 101, pp. 63–80, Jun. 2016.
- [24] M. M. Rathore, A. Paul, A. Ahmad, M. Anisetti, and G. Jeon, "Hadoop-based intelligent care system (HICS): Analytical approach for big data in IoT," *ACM Trans. Internet Technol.*, vol. 18, no. 1, p. 8, Dec. 2017.
- [25] B. N. Silva, M. Khan, C. Jung, J. Seo, Y. Yoon, J. Kim, S. Jin, J. Kang, and K. Han, "Planning of smart cities: Performance improvement using big data analytics approach," in *Proc. 4th Int. Conf. Adv. Comput., Electron. Commun. Inst. Res. Eng. Doctors*, 2016, pp. 51–55.
- [26] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of Things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.
- [27] R. Tönjes, M. I. Ali, P. Barnaghi, S. Ganea, F. Ganz, M. Haushwirth, B. Kjærgaard, D. Kümper, A. Mileo, S. Nechifor, A. Sheth, V. Tsiatsis, and L. Vestergaard, "Real time iot stream processing and large-scale data analytics for smart city applications," in *Proc. Eur. Conf. Netw. Commun.*, 2014, pp. 1–5.
- [28] B. Cheng, S. Longo, F. Cirillo, M. Bauer, and E. Kovacs, "Building a big data platform for smart cities: Experience and lessons from santander," in *Proc. IEEE Int. Congr. Big Data*, Jun. 2015, pp. 592–599.
- [29] M. M. Rathore, A. Paul, A. Ahmad, and G. Jeon, "IoT-based big data: From smart city towards next generation super city planning," *Int. J. Semantic Web Inf. Syst.*, vol. 13, no. 1, pp. 28–47, 2017.
- [30] V. G. Menon and J. Prathap, "Vehicular fog computing: Challenges applications and future directions," *Int. J. Veh. Telematics Inf. Syst.*, vol. 1, no. 2, pp. 15–23, 2017.
- [31] M. M. Rathore, A. Ahmad, A. Paul, and G. Jeon, "Efficient graph-oriented smart transportation using Internet of Things generated big data," in *Proc. 11th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2015, pp. 512–519.
- [32] Z. Shelby, K. Hartke, and C. Bormann, *The Constrained Application Protocol (CoAP)*, document RFC 7252, 2014.
- [33] K. Kuladinithi, O. Bergmann, T. Pötsch, M. Becker, and C. Görg, "Implementation of coap and its application in transport logistics," in *Proc. IP+SN*, Chicago, IL, USA, 2011, pp. 1–6.
- [34] A. Ludovici, P. Moreno, and A. Calveras, "TinyCoAP: A novel constrained application protocol (CoAP) implementation for embedding RESTful Web services in wireless sensor networks based on TinyOS," *J. Sens. Actuator Netw.*, vol. 2, no. 2, pp. 288–315, May 2013.



HUI ZHANG received the B.Sc. degree in communication engineering from Henan Polytechnic University, China, in 2007, the M.Sc. degree from the School of Energy Science and Engineering, Henan Polytechnic University, in 2010, and the Ph.D. degree from the School of Energy and Mining Engineering, China University of Mining and Technology, in 2013. He is currently an Associate Professor with the School of Energy Science and Engineering, Henan Polytechnic University.

He has authored several articles in journal and conferences, and holds several patents. His research interests include mining communication and smart mine.



MUHAMMAD BABAR received the bachelor's degree (Hons.) in computer sciences from the University of Peshawar, Pakistan, in 2008, and the Master of Science and Ph.D. degrees in computer software engineering from National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2012. He is currently with Iqra University, Islamabad. He has published his research work in various IEEE and ACM/Springer international conferences and journals. His research interests include big data analytics, the Internet of Things (IoT), smart city design and planning, and Social Web of Things (SWOT). He is an active Reviewer and a guest editor in the reputed journals.



MUHAMMAD USMAN TARIQ received the bachelor's and Master of Science degrees in computing, with a specialization in software engineering, and the Ph.D. degree (Hons.) in management from Calsouthern, USA. He has more than 13 years' experience in industry and academia. He has a passion for learning and development, project management, and training that made him achieve four patents. His research interests include management, the IoT, six sigma, knowledge management, information technology, economics, organizational change, facial recognition, biomedical devices, and computer science.



MIAN AHMAD JAN received the Ph.D. degree in computer systems from University of Technology Sydney (UTS), Australia, in 2016. He is currently a Researcher with Ton Duc Thang University, Vietnam. His research has been published in various prestigious the IEEE TRANSACTIONS and Elsevier Journals. His research interests include security and privacy in the Internet of Things, and wireless sensor networks. He was a recipient of various prestigious scholarship during his studies, notably the International Research Scholarship (IRS) at the UTS, and the Commonwealth Scientific Industrial Research Organization (CSIRO) scholarships. He has been received the Best Researcher awarded for the year 2014 at the UTS, Australia. He has been the general Co-Chair of Springer/EAI 2nd International Conference on Future Intelligent Vehicular Technologies, in 2017. He has been a guest editor of numerous special issues in various prestigious journals, such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, *Future Generation Computer Systems* (Elsevier), *Mobile Networks and Applications* (MONET) (Springer), *Ad Hoc & Sensor Wireless Networks*, and *MDPI Information*.



VARUN G. MENON (Senior Member, IEEE) is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include the Internet of Things, fog computing and networking, underwater acoustic sensor networks, cyber psychology, hijacked journals, ad-hoc networks, and wireless sensor networks. He is a Distinguished Speaker of ACM Distinguished Speaker. He is currently a Guest Editor of the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, the IEEE SENSORS JOURNAL, the *IEEE Internet of Things Magazine*, and the *Journal of Supercomputing*. He is an Associate Editor of *IET Quantum Communications*. He is also an Editorial Board Member of the IEEE Future Directions: Technology Policy and Ethics.








XINGWANG LI (Senior Member, IEEE) received the B.Sc. degree from Henan Polytechnic University, Jiaozuo, China, in 2007, the M.Sc. degree from the University of Electronic Science and Technology of China, in 2010, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, in 2015.

From 2010 to 2012, he was working as an Engineer with Comba Telecom Ltd., Guangzhou, China. From 2016 to 2018, he was also a Visiting

Scholar with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. From 2017 to 2018, he was a Visiting Scholar with Queen's University Belfast, Belfast, U.K. He is currently an Associate Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University. His research interests include MIMO communication, cooperative communication, hardware constrained communication, non-orthogonal multiple access, physical layer security, unmanned aerial vehicles, and the Internet-of-Things. He has served as many TPC members, such as the IEEE/CIC International Conference on Communications in China (ICCC'2019) and the IEEE Global Communications Conference 2018 (Globecom'18). He is also an Editor on the Editorial Board of IEEE ACCESS, *Computer Communications*, and *KSI Transactions on Internet and Information Systems*. He is also the Lead Guest Editor for the Special Issue on Recent Advances in Physical Layer Technologies for the 5G-Enabled Internet of Things of Wireless Communications and Mobile Computing and the Lead Guest Editor for the Special Issue on Recent Advances in Multiple Access for 5G-enabled IoT.

...

Lightweight Mutual Authentication and Privacy-Preservation Scheme for Intelligent Wearable Devices in Industrial-CPS

Mian Ahmad Jan, *Senior Member, IEEE*, Fazlullah Khan , *Senior Member, IEEE*,
 Rahim Khan , *Member, IEEE*, Spyridon Mastorakis , *Member, IEEE*,
 Varun G. Menon , *Senior Member, IEEE*, Mamoun Alazab, *Senior Member, IEEE*,
 and Paul Watters , *Senior Member, IEEE*

Abstract—Industry 5.0 is the digitalization, automation, and data exchange of industrial processes that involve artificial intelligence, industrial Internet of Things (IIoT), and industrial cyber-physical systems (I-CPS). In healthcare, I-CPS enables the intelligent wearable devices to gather data from the real-world and transmit to the virtual world for decision-making. I-CPS makes our lives comfortable with the emergence of innovative healthcare applications. Similar to any other IIoT paradigm, I-CPS capable healthcare applications face numerous challenging issues. The resource-constrained nature of wearable devices and their inability to support complex security mechanisms provide an ideal platform to malevolent entities for launching attacks. To preserve the privacy of wearable devices and their data in an I-CPS environment, in this article we propose a lightweight mutual authentication scheme. Our scheme is based on client-server interaction model that uses symmetric encryption for establishing secured sessions among the communicating entities. After mutual authentication,

the privacy risk associated with a patient data is predicted using an AI-enabled hidden Markov model. We analyzed the robustness and security of our scheme using Burrows–Abadi–Needham logic. This analysis shows that the use of lightweight security primitives for the exchange of session keys makes the proposed scheme highly resilient in terms of security, efficiency, and robustness. Finally, the proposed scheme incurs nominal overhead in terms of processing, communication and storage and is capable to combat a wide range of adversarial threats.

Index Terms—Artificial intelligence (AI), authentication, client-server model, industrial cyber-physical systems (I-CPS), Industrial Internet of Things (IIoT), privacy, security.

I. INTRODUCTION

THE latest developments in Industry 5.0 have enabled the integration of industrial Internet of Things (IIoT), industrial cyber-physical systems (I-CPS), big data technologies, cloud computing, and artificial intelligence (AI) [1]. It has resulted in collecting huge amounts of data from different industrial applications using intelligent IIoT devices. For example, in I-CPS enabled healthcare applications, wearable devices implanted on a patient body are capable to stream the real-time data to the cyberspace for computation, storage, and bigdata analytics [2]. I-CPS facilitate the healthcare entities with cyber computational capabilities for making quicker decisions. To deliver high-quality services at low cost, the healthcare practitioners need to adopt I-CPS based practices. In a healthcare ecosystem, the smart devices of IIoT are capable to gather, analyze, and broadcast a diverse range of data. These devices ensure the real-time monitoring of patients to save lives in an event of emergency, e.g., heart failure, severe pain, asthma, etc. The proliferation in mobile communication bridges the gap among these smart devices and the practitioners by providing seamless and reliable delivery of gathered data [3]. The patient-centric approach of I-CPS enables the remote monitoring of patients with shorter hospital stays and, in most cases, avoiding the hospital altogether. Using industrial techniques in I-CPS, we need to consider the patients’ willingness and feelings about these techniques.

Manuscript received August 21, 2020; revised November 5, 2020; accepted December 1, 2020. Date of publication December 10, 2020; date of current version May 3, 2021. This work was supported by a pilot award from the Center for Research in Human Movement Variability and the NIH under Grant P20GM109090 and a planning award from the Collaboration Initiative of the University of Nebraska system. Paper no. TII-20-4001. (*Corresponding author: Fazlullah Khan.*)

Mian Ahmad Jan and Rahim Khan are with the Department of Computer Science, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan.

Fazlullah Khan is with the Institute of Social and Economic Research, Duy Tan University, Da Nang 550000, Vietnam, and also with the Department of Computer Science, Abdul Wali Khan University Mardan, Mardan 23200, Pakistan (e-mail: fazlullahkhan@duytan.edu.vn).

Spyridon Mastorakis is with the Computer Science Department, University of Nebraska Omaha, Omaha, NE 68 182-0002 USA.

Varun G. Menon is with the Computer Science Engineering Department, SCMS Group of Educational Institutions, Ernakulam 683576, India.

Mamoun Alazab is with the Charles Darwin University, Casuarina, NT 0811, Australia.

Paul Watters is with the School of Engineering and Mathematical Sciences, La Trobe University Melbourne, Melbourne, VIC 3086, Australia.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2020.3043802>.

Digital Object Identifier 10.1109/TII.2020.3043802

The increasing use of industrial techniques in I-CPS brings new risks, vulnerabilities, and challenges for practitioners and their patients. Not only the IIoT devices and their data, but the complete healthcare ecosystem needs to be secured against the adversarial attacks [4]. The IIoT devices hosting the healthcare applications contain sensitive information, e.g., date of birth, prescriptions, medical histories, and social security numbers of the patients. These devices act as gateways to the secured Internet. An adversary may compromise these devices to inject fabricated data, ransomwares and other malwares into the network [5]. In the traditional computing platforms, cybersecurity is a matured domain and can defend against most of these adversarial threats. The existing cybersecurity solutions include cryptographic techniques, secured protocols and privacy protections that require ample of network resources. However, the security requirements and system architecture of IIoT-based I-CPS are different and as such, these existing solutions are not directly applicable [6], [7]. In I-CPS, most of the devices are connected to the Internet for the first time. It is extremely difficult to predict the nature of adversarial threats posed by these devices, if compromised. To secure the I-CPS, data integrity, data confidentiality, data availability, authenticity, and nonrepudiation need to be in place [8].

I-CPS enabled healthcare applications consist of resource-constrained sensor nodes and requires lightweight and low-cost protective measures. To deal with the aforementioned challenges, datagram transport layer security (DTLS) is proposed as a lightweight secured approach for these applications of I-CPS [9]. In literature, numerous DTLS-enabled authentication approaches exist for secured data transmission, and privacy of patients in healthcare applications [8], [10]–[12]. In [10], the authors proposed an end-to-end authentication scheme for a mobility-enabled healthcare application. A certificate-based DTLS handshake approach is used for the end-users authentication and authorization. The proposed scheme provides robust mobility using the interconnected smart gateways at the expense of computational overhead due to the use of certificate-based DTLS. In [11], a secured authentication approach was proposed using a body sensor network. The use of crypto-primitives enable the proposed approach to achieve system efficiency and robustness, and at the same time, provides the transmission confidentiality and authentication among the wearables and a backend server. However, the use of an asymmetric algorithm, i.e., Elliptic-curve cryptography, incurs additional overhead for these intelligent wearables. In [12], the authors presented a lightweight DTLS-enabled authentication approach for wearables of a smart healthcare system. The proposed approach allows a user to authenticate his/her wearable device(s) and a mobile terminal, prior to establishing a session key among them. The use of bitwise exclusive-OR (XOR) and hash functions make the proposed scheme significantly lightweight for the resource-constrained wearables. The security analysis of DTLS via different techniques, such as the random oracle model [13] and the Burrows–Abadi–Needham (BAN) logic [14], showed that the use of DTLS for secured message exchanges leaves a handful of payload for most of healthcare applications. This

remaining payload is not sufficient for these applications due to their larger packet sizes, e.g., healthcare streaming applications.

Besides authentication, the privacy of patients and their data needs to be dealt with utmost care in I-CPS. Different machine learning (ML) algorithms have been used in the literature for this purpose. An ML-based privacy-preserved healthcare framework was presented in [15]. This framework uses ML-based scoring service for the classification, and cryptographic algorithms for data protection. It is a cloud-based framework for privacy risk prediction in healthcare applications. In [16], the authors have provided general guidelines about privacy challenges in AI-based healthcare applications. The proposed work mainly focuses on policies for the usage of AI-based healthcare guidelines to preserve the privacy of patients. In [17], the authors have proposed a framework known as ModelChain. This framework uses ML and blockchain for privacy preservation of patients in a decentralized environment. ModelChain embeds the intelligence in private blockchains to preserve the privacy of patients and increases the interoperability between healthcare centers. In [18], the authors have discussed AI-based cyber-physical security and privacy for healthcare applications. They proposed a ciphertext-policy attribute-based encryption scheme. In the proposed scheme, complex computation tasks are offloaded to the third parties for reducing load on wearables while preserving their privacy at the same time. Most of these approaches use asymmetric encryption that require ample resources on part of the wearables to perform effectively.

In view of the resource-constrained nature of the healthcare devices, we propose a lightweight mutual authentication scheme for I-CPS. The proposed scheme uses symmetric encryption for the exchange of handshake messages that can be used as an alternative to the DTLS scheme. We perform its security analysis using BAN logic to determine whether the exchanged information is trustworthy and secured against eavesdropping attack, and predict the privacy leakage using a hidden Markov model (HMM). The hidden and observable states of HMM are used to measure the risk of data leakage by preserving the privacy of a patient and his/her connected devices. The major contributions of the proposed work are as follows.

- 1) An authentication scenario is proposed in which a client-server authentication takes place only if the clients, i.e., wearable patients, are within the coverage of their designated servers. Each server maintains a record of preshared keys for the clients in its proximity. For the aforementioned scenario, a set of theorems are proposed and their proofs are provided. Each theorem corresponds to a handshake message that takes into account the possibility and probability of an adversarial attack.
- 2) A privacy risk prediction model is proposed using HMM. The proposed model is used to predict the risk of privacy leakage of the patient identity and his/her data. If the privacy risk is predicted, the patients' data is altered with a loss in utility. To the best of our knowledge, this is the first ever work on HMM for predicting the privacy leakage.
- 3) Security analysis of mutual authentication and session key exchange of our proposed scheme is performed using

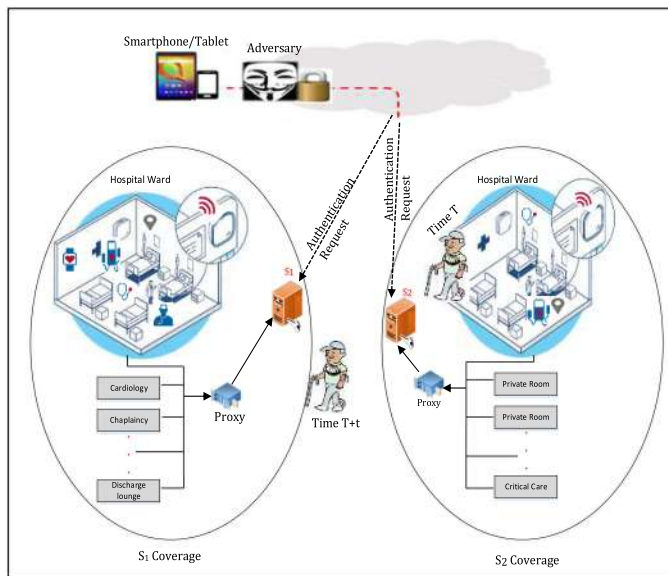


Fig. 1. Smart and secured healthcare facilitation center.

BAN logic. The security goals are set according to the exchanged messages and are proven using the postulates of BAN logic.

The rest of the article is organized as follows. In Section II, the network and threat model is briefly explained. In Section III, our proposed lightweight mutual authentication and privacy-preservation scheme are presented. In Section IV, the security analysis of the proposed scheme is performed using BAN logic. In Section V, we present the experimental results of our proposed scheme. Finally, the article is concluded in Section VI.

II. NETWORK AND THREAT MODEL

We have considered a healthcare facilitation center, i.e., a hospital within an industry, as a case study of our proposed I-CPS scheme. Various units such as critical care, chaplaincy, cardiology, radiology, wards, and discharge lounge, along with private rooms provide timely healthcare facilities to the patients. These units and rooms are connected to remote servers for storing the patients’ data and other credentials to provide on-demand and responsive services. In Fig. 1, the sensor-embedded wearables, i.e., clients, in various units and rooms are connected to servers via their proxies. Each server facilitates a number of clients within its coverage region. A client is static in the context of the server’s coverage region, i.e., a client remains within the coverage region of its associated server. For seamless and interoperable communication, these clients need to establish secured communication links to their concerned servers.

In healthcare applications, any adversarial attack can lead to the loss of precious lives and the associated medical data. An adversary may establish secured connections to the servers if its authentication requests are accepted. The smart healthcare environment of Fig. 1 is prone to various types of adversarial attacks. An adversary may infiltrate the network by seizing

the identities of clients and servers to pose various threats. It is important to mention that in Fig. 1 the adversary uses a smartphone to launch the attacks. Moreover, it may clone itself for a large-scale adversarial effect on the overall system. To prevent such threats, we propose a lightweight mutual authentication approach for resource-starving intelligent wearables. Our authentication approach is resilient against the following threats.

- 1) Replay: An adversary may replay a stream of previously transmitted messages to the clients or servers.
- 2) Forward and backward secrecy: An adversary may launch this attack by seizing the session key to predict the outcome of previous or future sessions.
- 3) Client and server impersonation: An adversary may impersonate a legitimate client to the server by fabricating the preshared key of the given client. Moreover, it may impersonate a legitimate server to one or more clients by fabricating the session key of the given server.
- 4) Anonymity and untraceability: An adversary may launch this attack by extracting the one-time nonces, and the identities of clients and servers from exchanged messages. In doing so, it may interlink various sessions to maliciously affect the clients and servers.
- 5) Eavesdropping: An adversary may launch active or passive eavesdropping by listening to the communication in transit. It may seize various messages, manipulate them, and may launch other types of attacks. The use of pseudorandom nonces in our approach restricts an adversary from launching this attack.
- 6) Denial of service (DoS): An adversary may broadcast excessive requests to the clients or servers to authenticate itself. By doing so, it may deprive the legitimate clients from exchanging their data with the legitimate servers. The use of preshared keys restricts an adversary from launching a DoS attack in our approach.

III. LIGHTWEIGHT MUTUAL AUTHENTICATION AND PRIVACY-PRESERVATION SCHEME

In this section, we discuss our mutual authentication and privacy preservation scheme for the healthcare facilitation center of Fig. 1. Numerous wearables within the hospital communicate with their concerned servers for authentication, as shown in Fig. 2. In this figure, A is the set of attackers, C is the set of clients, and S is the set of servers, where C_i can communicate either directly with S_j or via a proxy (P). Our proposed scheme comprises of C_i clients and S_j servers, where $i = \{1, 2, 3, \dots, I\}$ and $j = \{1, 2, 3, \dots, J\}$, such that $i, j \in N$, and $i > j$. Here, N is the total number of C_i and S_j in the network, i.e., $N = C_i \cup S_j$. C_i are dynamic in nature and may change their positions quite frequently, whereas S_j are static in nature. Our proposed scheme initiates a four-way handshake between any C_i and S_j for mutual authentication. If the handshake is successful, S_j provides a session key to C_i for data transmission. The list of Symbol used in authentication is given in Table I. We discuss mutual authentication in Section III-A and privacy risk prediction using HMM in Section III-B.

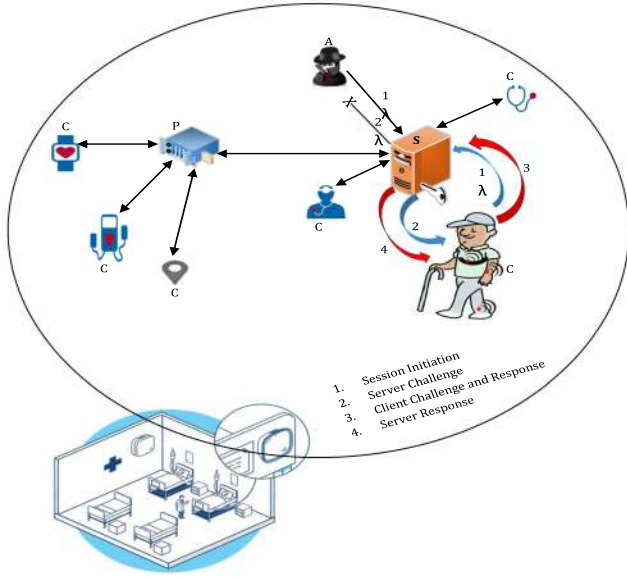


Fig. 2. Proposed mutual authentication scheme.

TABLE I
SYMBOLS AND MEANING USED IN AUTHENTICATION

Symbols	Meanings
C_i	Client i
S_j	Server j
A_k	adversary k
λ	128-bit pre-shared key
μ	128-bit session key
ID_i	Identity of Client i
$h()$	hash function
η	one-time 128-bit pseudo nonce
$\gamma_{challenge}$	256-bit server challenge
$\gamma_{response}$	Server response to client
$\beta_{challenge}$	256-bit client challenge

A. Mutual Authentication

Each C_i periodically collects the desired data and transmits to the nearest S_j . However, prior to the data transmission, both C_i and S_j need to be authenticated. Our lightweight authentication scheme verifies the identities of C_i and S_j before their engagement for data exchange. The authentication is performed using the following four handshake messages.

- 1) Session initiation.
- 2) Server challenge.
- 3) Client response and challenge.
- 4) Server response.

Initially, both C_i and S_j are assumed to be unauthentic, and thus, untrustworthy. Prior to mutual authentication, each C_i is assigned a unique 128-b preshared key (λ_i), and an identity (ID_i) in an offline phase. These secret primitives are also shared with their associated S_j , located in their vicinity. The offline phase is a prerequisite for the initialization of C_i and S_j , respectively. Next, each C_i initiates a session request to its associated S_j . This session initiation request contains the encrypted identity

$\lambda_i(ID_i)_{h()}$ of C_i , i.e., ID_i is encrypted by C_i using its λ_i and hashed using $h()$. The transmitted request message is meaningless to the neighboring C_{i-1} clients and adversaries A_k , where $k=\{1, 2, 3, \dots, K\}$, such that $k \notin \{i, j\}$. The recipient, be it C_i , S_j or A_k , needs to decrypt $\lambda_i(ID_i)_{h()}$ with the same λ_i and $h()$. Please note that the mode of wireless communication means that any device can intercept the session initiation request.

Theorem 1: At least one legitimate C_i , not an adversary A_k , initiates a session with the corresponding S_j .

Proof: Each C_i shares its λ_i with its associated S_j in an offline phase. The set of identities and keys of C_i , i.e., $\{ID_1, ID_2, ID_3, \dots, ID_i\}$ and $\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_i\}$, respectively, are stored by S_j in a database. An A_k may initiate a session request by transmitting a message $\lambda_k(ID_k)_{h(ID_k)}$, encrypted with a fabricated λ_k and $h(ID_k)$. S_j checks the authenticity of this request by retrieving the corresponding decrypting key λ_k . Since, $\lambda_k \notin \{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_i\}$, S_j assumes that the request was initiated by an adversary. The λ_i for encryption and decryption is computed using the equality to compute λ_i and λ'_i , respectively [19].

$$\lambda_i = \text{from-state} \oplus \text{Round}_{0-9} \oplus \text{Add-Round}_{key} \oplus \text{to-state } \lambda'_i \\ = \text{from-state} \oplus \text{Round}'_{0-9} \oplus \text{Add-Round}_{key} \oplus \text{to-state}$$

Here, λ_i and λ'_i are the secret encryption and decryption keys, where $\lambda_i = \lambda'_i$. The only difference is that in λ_i , from-state represents the plain text and to-state represents the cipher text. For λ'_i , from-state and to-state work oppositely to λ_i . Round is a function used to compute a unique key every time [19], as explained below.

$\text{Round}_0 \text{ state}_{key} = \text{AddRound}_{key} (\text{ShiftRows}(\text{SubBytes}(\text{state}))) \oplus (\text{Round}_{n+1} \text{ state}_{key} = \text{Round}_n \text{ state}_{key} (\text{AddRound}_{key} (\text{MixColumns} (\text{ShiftRows} (\text{SubBytes}(\text{state}))))))$. where, AddRound is a pairwise XOR operation, ShiftRows applies permutation to the block, SubBytes applies an S-Box operation on every state and MixColumns transforms every column of the metric.

The session initiation request is terminated by S_j either by ignoring it or by sending a denial message, i.e., when $\lambda_k \notin \{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_i\}$. Hence, any C_i with an appropriate λ_i is capable of initiating the session with an S_j . Conversely, if the session initiation request, encrypted with λ_i , is received by an A_k , the latter is unable to decrypt it. This is because the λ'_i is known only to encrypting C_i and to the associated S_j .

Upon the reception of a session initiation request, S_j retrieves $\lambda_i(ID_i)_{h()}$ and decrypts it with λ'_i and $h()$ to check ID_i in it. If the embedded ID_i matches with an entry in S_j database, it means that the session initiation request was received from a legitimate C_i . At this point, S_j creates a challenge for the concerned C_i to confirm its authenticity by establishing a session with it. For this purpose, S_j generates a 128-b session key (μ_j), and a temporary one-time 128-b pseudononce (η_{server}). The nonce is computed by generating two pseudorandom numbers η_{s1} and η_{s2} , and an XOR operation is performed on them using (1).

$$\eta_{server} = \eta_{s1} \oplus \eta_{s2}. \quad (1)$$

Next, an XOR operation is performed on μ_j and λ_i , and their 128-b resultant is concatenated with η_{server} . Finally, $\lambda_i \oplus \mu_j | \eta_{server}$ is encrypted with λ_i and hashed with $h()$ to generate a

256-b server challenge ($\gamma_{\text{challenge}}$) as shown in (2). The advanced encryption standard (AES) of 128 b is used for symmetric encryption in Cipher block chaining mode.

$$\gamma_{\text{challenge}} = \text{AES}((\lambda_i, (\lambda_i \oplus \mu_j | \eta_{\text{server}}))_{h()}). \quad (2)$$

Theorem 2: An encrypted $\gamma_{\text{challenge}}$ is resolved **iff** a C_i or an A_k has the required λ'_i for decryption.

Proof: Any C_i receiving the $\gamma_{\text{challenge}}$ that contains μ_j needs to have the required λ'_i for decryption. Assume that the $\gamma_{\text{challenge}}$ is received by A_k , and $f(x_k)$ is the function used by A_k to compute a matching λ_i from the set $\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_i\}$ as shown in (3).

$$f(x_k) = \{S_j, (C_1, \lambda_1), (C_2, \lambda_2), (C_3, \lambda_3), \dots, (C_i, \lambda_i)\}. \quad (3)$$

Here, $\{C_1, C_2, C_3, \dots, C_i\}$ represents the client devices' IDs that are generated by A_k based on historic data collection, and $\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_i\}$ are their dummy secret keys. These dummy keys are computed using (4).

$$\lambda_i = C_i \oplus \text{statistics}(\lambda_i). \quad (4)$$

Since, the $\gamma_{\text{challenge}}$ is encrypted with a particular λ_i known only to a legitimate C_i and S_j , A_k will compute and apply different λ_i values, as shown in (3), to decipher the cipher text of (2). However, the success probability is $\frac{1}{2^{128}}$. Thus, A_k will not be able to decrypt the $\gamma_{\text{challenge}}$ within a stipulated time. Conversely, if a C_i has the required λ_i , then it will decrypt $\gamma_{\text{challenge}}$ within its stipulated time. Hence, an encrypted $\gamma_{\text{challenge}}$ is resolved only by a single C_i that has the required λ_i .

Upon the reception of $\gamma_{\text{challenge}}$, if C_i successfully deciphers it, then it will have access to the corresponding η_{server} and μ_j . Additionally, it proves the authenticity of C_i to S_j . It is because η_{server} and μ_j are known only to a given S_j and λ_i to the concerned C_i . To authenticate an S_j , C_i generates a client challenge for the given S_j . Initially, a temporary one-time 128 b pseudononce (η_{client}) is computed by generating two pseudorandom numbers η_{c_1} and η_{c_2} . Next, an XOR operation is performed on them using (5).

$$\eta_{\text{client}} = \eta_{c_1} \oplus \eta_{c_2}. \quad (5)$$

Next, an XOR operation is performed on η_{server} and λ_i , their resultant is concatenated with η_{client} , and finally encrypted with μ_j to generate a 256-b client challenge $\beta_{\text{challenge}}$, as shown in (6).

$$\beta_{\text{challenge}} = \text{AES}((\mu_j, (\eta_{\text{server}} \oplus \lambda_i | \eta_{\text{client}}))_{h()}). \quad (6)$$

Theorem 3: An encrypted $\beta_{\text{challenge}}$ is resolved and responded **iff** a device, such as S_j , has the shared information, i.e., η_{server} and μ_j .

Proof: The μ_j and η_{server} are known only to a given C_i and S_j . Assume that an A_k receives $\beta_{\text{challenge}}$ and tries to decrypt it using a probabilistic function $g(x)$. This function is used to compute the desired μ_j by using (7).

$$g(x) = \text{probability}((C_1, \mu_1), (C_2, \mu_2), \dots, (C_i, \mu_j)). \quad (7)$$

The function $g(x)$ utilizes the C_i and S_j information to return a single pair of values for A_k , i.e., (ID_i, μ_j) . However, this

scenario is applicable only if A_k maintains a complete record of the overall communication between C_i and S_j , which is not a realistic assumption especially in a resource-constrained health-CPS environment. In addition to μ_j and λ_i values that are known only to C_i and S_j , A_k needs to verify its authenticity to C_i as well. Conversely, if $\beta_{\text{challenge}}$ is received correctly by the concerned S_j , then the latter deciphers $(\eta_{\text{server}} \oplus \lambda_i | \eta_{\text{client}})_{h()}$ of (6) correctly with μ_j and $h()$ to retrieve η_{client} . Thus, $\beta_{\text{challenge}}$ of a given C_i is resolved by a particular S_j that possesses the required μ_j .

Finally, during the server response, the concerned S_j creates a response by concatenating the C_i 's η_{client} to its μ_j , and generates an encrypted server response (γ_{response}) using λ_i (8).

$$\gamma_{\text{response}} = \text{AES}((\lambda_i, \{\eta_{\text{client}} | \mu_j\})_{h()}). \quad (8)$$

Upon reception, a C_i having a valid λ'_i will be able to decipher γ_{response} and retrieve η_{client} to confirm the authenticity of the given S_j .

Theorem 4: The encrypted γ_{response} of an S_j is decrypted by a C_i **iff** it has the required λ_i .

Proof: In the prerequisite offline phase, C_i shared their λ_i with their concerned S_j . The γ_{response} is decrypted by an A_k only if it has the required λ_i , which is not the case. An A_k uses the functions $f(x)$ and $g(x)$, as discussed earlier, to find an exact copy of λ_i .

$$\lambda_i = f(x) \oplus g(x). \quad (9)$$

Where, $f(x)$ and $g(x)$ return a pair of values, i.e., (C_i, λ_i) and (C_i, μ_j) , respectively. However by adopting the approach of (9), A_k will only be able to obtain μ_j at the expense of excessive resource consumption. However, it will still not be able to collect the desired λ'_i that is required to decrypt γ_{response} . Conversely, if γ_{response} is received by the concerned C_i having the appropriate λ_i , it will be able to decrypt this message within the stipulated time. Thus, a given C_i having λ_i is able to successfully decrypt the γ_{response} of C_i . Upon successful decryption of γ_{response} , both C_i and S_j have mutually authenticated each other and are authorized to exchange data. After successful authentication, data are transmitted from C_i to S_j . During data transmission and storage at S_j , the C_i privacy can be leaked, and hence needs to be preserved. To solve the privacy leakage issues, we use HMM to predict the privacy of C_i . In the next section, we present an approach to predict the privacy risks of C_i using HMM.

B. Privacy Risk Prediction Using HMM

In this section, we predict the risk of a client's privacy leakage using HMM. In HMM, states are partially observed that helps in solving real-world problems using sequential or temporal data. The aim of the proposed model is to measure the risk of data privacy leakage using HMM. The graphical representation of HMM is shown in Fig. 3. The HMM uses two sets of random variables, hidden variable $\mathbf{H} = \{H_1, H_2, \dots, H_m\}$ and observed variable $\mathbf{O} = \{O_1, O_2, \dots, O_n\}$, where $\mathbf{O} \in \{\text{discrete values, real values, } R^d\}$. In our proposed scheme, \mathbf{H} is the data generated by the patients and \mathbf{O} is the usage pattern of C_i devices associated

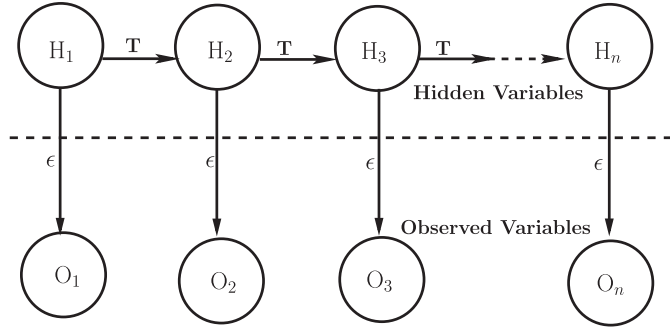


Fig. 3. Graphical demonstration of HMM.

with a patient. The joint probability distribution of HMM in terms of \mathbf{H} and \mathbf{O} is given in (10).

$$P(O_1, O_2, \dots, O_n, H_1, H_2, \dots, H_n) \\ = P(H_1)P(O_1|H_1) \prod_{k=2}^n P(H_k|H_{k-1})P(O_k|H_k). \quad (10)$$

1) **Probabilities of the HMM:** The HMM works on the initial probability $\pi(i)$, the observation probability $E_i(O)$, and the transition probability T_{ij} . The initial probability ($\pi(i)$) of a patient's data in the context of HMM is given in (11),

$$\pi(i) = P(H_1 = i), \text{ for } i \in \{1, 2, \dots, m\}. \quad (11)$$

where, $\pi(i)$ is based on the previous data shared by a patient, which include personal identification (PI) such as patient's name, patient's location, and his/her illness, etc. $\pi(i)$ is important in the privacy risk identification because it reveals PI of a patient that can be linked to anonymized data shared by the patient using HMM. The initial risk probability of a client C_i is computed by observing data D_t . (11) can be re-written as

$$\pi(C_i) = \begin{cases} p(C_i|D_t) > 0, & \text{for a patient having a shared PI} \\ p(C_i|D_t) = 0, & \text{for a patient having no shared PI} \end{cases} \quad (12)$$

E_i is the probability distribution on \mathbf{O} , which can be defined as a probability density function for $\{H_1, H_2, \dots, H_m\}$ and $\forall O \in \mathbf{O}$, it can be written as

$$E_i(O) = P(O|H_k = i), \text{ for } i \in \{1, \dots, m\}, \text{ and } O \in \mathbf{O}. \quad (13)$$

When \mathbf{O} takes discrete random numbers, then (13) can be written as the probability mass function as shown in (14).

$$E_i(O) = P(O_k = O|H_k = i), \text{ for } i \in \{1, \dots, m\}, \\ \text{and } O \in \mathbf{O}. \quad (14)$$

E_i is the probability of the data stored previously by C_i that can reveal the consistency in the patient data and his/her usage pattern. We modeled E_i as the probability of data (D_t) shared by various patients in (12). It is needed to embed inconsistency in the frequency of data sharing by a patient. The data frequently shared by a patient reveal his/her concern of causing higher risk, that can easily be inferred from the shared data. To increase the

inconsistency in the patient data and reduce the privacy risk, a weight is multiplied with each probability and then it is inverted, as shown in (15).

$$W_{E_i(O)} = 1 - \frac{p(C_i|D_t)}{\text{count}(C_i|D_t)}. \quad (15)$$

where, $1/\text{count}(C_i|D_t)$ is the weight multiplied to each probability.

The transition probability T_{ij} is given in (16), which is the conditional probability of current data given a sequence of previously shared data.

$$T_{ij} = P(H_{k+1} = j|H_k = i), \forall i, j \in \{1, 2, \dots, m\}. \quad (16)$$

Eq. (16) models the distinctiveness of a patient's data from all other patients because the data distinguishability depends on the previous data. The T_{ij} between $p(O_j|O_{j-1})$ are weighted by the number of occurring transitions. To decrease the distinctiveness and privacy risk in the patient data, weighted transition probabilities are computed as in (17).

$$W_{T_{ij}} = \frac{p(O_j|O_{j-1})}{\text{count}(O_j|O_{j-1})}. \quad (17)$$

where, $1/\text{count}(O_j|O_{j-1})$ is the weight multiplied to each probability.

The probability of a patient's (C_i) privacy along with a sequence of his/her observed data $O_1 \rightarrow O_2 \rightarrow \dots \rightarrow O_j$ is calculated based on the Markov probability of (10),

$$p(O_1, \dots, O_j|C_i) = \min(\text{HMM}_{PI|C_i}) \times \omega_T \\ \times p(O_1) \times (1 - \omega_O \times p(C_i|O_1)) \\ \times \prod_{k=2}^n \omega_T \times p(O_k|O_{k-1}) \times (1 - \omega_O \times p(C_i|O_k)) \quad (18)$$

where, ω_T is $1/\text{count}(O_j|O_{j-1})$, and ω_O is $1/\text{count}(C_i|D_t)$. The $\text{HMM}_{PI|C_i}$ returns the list of privacy probabilities computed from the PI. It includes probabilities from the paths where E_i of a patient is greater than 0.

Upon identification of the privacy risk using (18), we alter the data to circumvent the privacy risk with a utility loss (ul). The ul uses a semantic similarity function [20], [21] to distinguish the original data D_t from the altered data D'_t , which is calculated as

$$ul(D, D') = 1.0 - \text{sim}(D, D') \quad (19)$$

The similarity function (sim) returns values within the range $[0, 1]$. The higher the similarity is, the lower ul is by using altered data. In this fashion, using HMM, the privacy of C_i is preserved. After privacy preservation, we need to analyze the correctness and efficiency of our proposed scheme. In the next section, we perform the security analysis of the proposed scheme using BAN logic.

IV. SECURITY ANALYSIS

In this section, we analyze the mutual authentication and session key (μ) of our proposed scheme using BAN logic [22]. BAN logic describes the trust of two parties involved in the

TABLE II
NOTATIONS AND RULES USED IN BAN LOGIC

Notations	Meanings
$P \equiv X$	P believes X
$P \triangleleft X$	P sees X or P receives X
$P \sim X$	P once said X
$P \Rightarrow X$	P has jurisdiction over X
$\#(X)$	X is fresh
$P \xleftrightarrow{K} Q$	P and Q may use the shared key K
$(X)_K$	X hashed under the key K
$\{X\}_\lambda$	X encrypted under the key K
Rule-1	Message meaning rule
Rule-2	Nonce verification rule
Rule-3	Jurisdiction rule
Rule-4	Freshness concatenation rule

communication. The notations and rules used in BAN logic are given in Table II.

1. The Postulates of BAN logic are given below,

1) Postulate of Rule-1 is,

$$\lambda_i = \frac{S_j \text{ believes } C_i \xleftrightarrow{\lambda_i} S_j, S_j \text{ sees } \{X\}_{\lambda_i}}{S_j \text{ believes } C_i \text{ said } X}$$

2) Postulate of Rule-2 is,

$$\frac{S_j \text{ believes fresh } (X), S_j \text{ believes } C_i \text{ said } X}{S_j \text{ believes } C_i \text{ believes } X}$$

3) Postulate of Rule-3 is,

$$\frac{S_j \text{ believes } C_i \text{ controls } X, S_j \text{ believes } C_i \text{ believes } X}{S_j \text{ believes } X}$$

4) Postulate of Rule-4 is,

$$\frac{S_j \text{ believes fresh } (X)}{S_j \text{ believes fresh } (X, Y)}$$

2. The following security goals must be met by the proposed scheme,

$$G_1. S_j | \equiv C_i \Rightarrow \mu_j$$

$$G_2. C_i | \equiv S_j \Rightarrow \mu_j$$

$$G_3. S_j | \equiv C_i | \equiv S_j \xleftrightarrow{\mu_j} C_i$$

$$G_4. C_i | \equiv S_j | \equiv C_i \xleftrightarrow{\mu_j} S_j$$

3. The proposed scheme should be transformed into an idealized form as below,

$$\text{Msg}_1. C_i \rightarrow S_j : (\lambda_i(\text{ID}_i))_{h()}$$

$$\text{Msg}_2. S_j \rightarrow C_i : (\lambda_i(\lambda_i \oplus \mu_j || \eta_{\text{server}}))_{h()}$$

$$\text{Msg}_3. C_i \rightarrow S_j : (\mu_j(\eta_{\text{server}} \oplus \lambda_i || \eta_{\text{client}}))_{h()}$$

$$\text{Msg}_4. S_j \rightarrow C_i : (\lambda_i(\eta_{\text{client}} || \mu_j))_{h()}$$

4. The following assumptions are mandatory for BAN logic.

$$A_1. C_i | \equiv S_j \xleftrightarrow{\lambda_i} C_i$$

$$A_2. S_j | \equiv C_i \xleftrightarrow{\lambda_i(\text{ID}_i)} S_j$$

$$A_3. C_i | \equiv S_j \xleftrightarrow{\lambda_i(\eta_{\text{server}})} C_i$$

$$A_4. S_j | \equiv C_i \xleftrightarrow{\lambda_i(\eta_{\text{client}})} S_j$$

$$A_5. C_i | \equiv \#(\lambda_i(\eta_{\text{server}}))$$

$$A_6. S_j | \equiv \#(\lambda_i(\eta_{\text{client}}))$$

$$A_7. S_j | \equiv C_i \Rightarrow S_j \xleftrightarrow{\mu_j} C_i$$

$$A_8. C_i | \equiv S^{\lambda_i} C_i \xleftrightarrow{\mu_j} S_j$$

5. We analyze security of the proposed scheme based on the idealized form,

$$s_1. \text{ From Msg}_1, \text{ we obtain } S_j \triangleleft (\lambda_i(\text{ID}_i))_{h()}$$

$$s_2. \text{ Applying Rule-1 and } A_2, \text{ we get } S_j | \equiv C_i | \sim (\lambda_i(\text{ID}_i))_{h()}$$

$$s_3. \text{ Applying Rule-4 and } A_6, \text{ we obtain } S_j | \equiv \#((\lambda_i(\text{ID}_i))_{h()})$$

$$\text{Then, we apply Rule-2 to get } S_j | \equiv C_i | \equiv \#(\lambda_i(\text{ID}_i))_{h()}$$

$$s_4. \text{ From Msg}_2, \text{ we obtain } C_i \triangleleft (\lambda_i(\lambda_i \oplus \mu_j || \eta_{\text{server}}))_{h()}$$

$$s_5. \text{ Applying Rule-1 and } A_1, \text{ we get } C_i | \equiv S_j | \sim (\lambda_i(\lambda_i \oplus \mu_j || \eta_{\text{server}}))_{h()}$$

$$s_6. \text{ Applying Rule-4 and } A_5, \text{ we obtain } C_i | \equiv \#(\lambda_i(\lambda_i \oplus \mu_j || \eta_{\text{server}}))_{h()}$$

$$\text{Then, we apply Rule-2 to get } C_i | \equiv S_j | \equiv (\lambda_i(\lambda_i \oplus \mu_j || \eta_{\text{server}}))_{h()}$$

$$s_7. \text{ From Msg}_3, \text{ we obtain } S_j \triangleleft (\mu_j(\eta_{\text{server}} \oplus \lambda_i || \eta_{\text{client}}))_{h()}$$

$$s_8. \text{ Applying Rule-1 and } A_2, \text{ we get } S_j | \equiv C_i | \sim (\mu_j(\eta_{\text{server}} \oplus \lambda_i || \eta_{\text{client}}))_{h()}$$

$$s_9. \text{ Applying Rule-4 and } A_6, \text{ we obtain } S_j | \equiv \#(\mu_j(\eta_{\text{server}} \oplus \lambda_i || \eta_{\text{client}}))_{h()}$$

$$\text{Then, we apply Rule-2 to get } S_j | \equiv C_i | \equiv (\mu_j(\eta_{\text{server}} \oplus \lambda_i || \eta_{\text{client}}))_{h()}$$

$$s_{10}. \text{ From Msg}_4, \text{ we obtain } C_i \triangleleft (\lambda_i(\eta_{\text{client}} || \mu_j))_{h()}$$

$$s_{11}. \text{ Applying Rule-1 and } A_1, \text{ we get } C_i | \equiv S_j | \sim (\lambda_i(\eta_{\text{client}} || \mu_j))_{h()}$$

$$s_{12}. \text{ Applying Rule-4 and } A_5, \text{ we obtain } C_i | \equiv \#(\lambda_i(\eta_{\text{client}} || \mu_j))_{h()}$$

$$\text{Then, we apply Rule-2 to get } C_i | \equiv S_j | \equiv (\mu_j(\eta_{\text{server}} \oplus \lambda_i || \eta_{\text{client}}))_{h()}$$

s₁₃. Applying the logic rule of BAN to s₁₂ and A₄, which split conjunctions that yields $C_i | \equiv S_j | \equiv C_i \xleftrightarrow{\mu_j} S_j$, (Goal 4)

s₁₄. Applying the logic rule of BAN to s₉ and A₃, which split conjunctions that yields $S_j | \equiv C_i | \equiv S_j \xleftrightarrow{\mu_j} C_i$, (Goal 3)

s₁₅. Applying Rule-3 to s₁₃ and A₈, which results in $C_i | \equiv S_j \Rightarrow \mu_j$, (Goal 2)

s₁₆. Applying Rule-3 to s₁₄ and A₇, which results in $S_j | \equiv C_i \Rightarrow \mu_j$, (Goal 1)

By performing the security analysis of our proposed scheme using BAN logic, the four security goals G₁, G₂, G₃, and G₄ are achieved. In the next section, we present the experimental results of our proposed scheme.

V. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our approach against existing state of the art schemes. For authentication, we used Netduino Plus 2 boards as clients and Netduino 3 boards as servers. The Netduino 3 boards were interfaced with MATLAB ThingSpeak server via the μ PLibrary 1.8.¹ This library abstracts the ThingSpeak API and works with these boards using NET Micro Framework. For privacy-preservation, we relied on MATLAB simulation at the ThingSpeak server. We evaluate the performance of our approach in term of computational, communication, storage overheads, and its resilience against various adversarial threats. These boards are resource-constrained and as such, lightweight authentication approaches need to be designed. For this purpose, we tested our proposed authentication in terms of computation, communication, and storage overhead incur by our authentication. For privacy preservation, we tested our approach through privacy risk prediction and privacy risk alleviation.

¹[Online]. Available: <https://www.nuget.org/packages/uPLibrary>

TABLE III
COMPUTATIONAL OVERHEAD COMPARISON

Schemes	Client Side	Server Side	Total Cost
<i>Li et. al [23]</i>	$3T_h+7T_{XOR}$	$4T_h+12T_{XOR}$	$7T_h+19T_{XOR}$
<i>Gupta et. al [6]</i>	$4T_h+4T_{XOR}$	$5T_h+3T_{XOR}$	$9T_h+7T_{XOR}$
<i>Gope et. al [7]</i>	$3T_h+1T_{XOR}$	$9T_h+4T_{XOR}$	$12T_h+5T_{XOR}$
<i>Chang et. al [9]</i>	$5T_h+4T_{XOR}$	$8T_h+1T_{XOR}$	$13T_h+5T_{XOR}$
<i>Proposed Scheme</i>	$2T_h+2T_{XOR}$	$2T_h+2T_{XOR}$	$4T_h+4T_{XOR}$

TABLE IV
COMMUNICATION OVERHEAD COMPARISON

Schemes	Number of messages	Number of bits
<i>Li et. al [23]</i>	4	4672
<i>Gupta et. al [6]</i>	5	3808
<i>Gope et. al [7]</i>	4	3184
<i>Chang et. al [9]</i>	4	3104
<i>Proposed Scheme</i>	4	896

In **Table III**, we provide a summary of the computational overhead analysis. We compare the execution time of our scheme against the existing schemes. In this table, T_h and T_{XOR} refer to the computational time needed to perform the hash and XOR operations. In our scheme, the encryption with λ_i and μ_j works similar to hashing. The proposed scheme requires only $2T_h+2T_{XOR}$ execution time at the C_i and S_j . The low computational overhead is contributed mainly to the lightweight mechanism adopted by $\lambda_k(ID_k)_{h()}$, $\gamma_{challenge}$, $\beta_{challenge}$, and $\gamma_{response}$ of the proposed scheme.

In **Table IV**, we provide a summary of the communication overhead analysis of our scheme against the existing schemes. The proposed scheme requires four handshake messages for the authentication. In this scenario, $\lambda_k(ID_k)_{h()}$ is 128 b, and $\gamma_{challenge}$, $\beta_{challenge}$, and $\gamma_{response}$ are 256 b each. Hence, total of 896 b communication overhead is incurred by these messages. In comparison, the existing schemes have much higher communication overhead due to the complex cipher-suites and the involvement of resource-intensive operators.

In **Table V**, we compare the storage overhead incurred by C_i and S_j of our proposed authentication scheme. In the proposed scheme, each C_i stores its ID_i and λ_i , respectively. On the other hand, each S_j stores ID_i and λ_i for n clients associated with it. In comparison, in [23] and [9], each C_i stores its ID_i and λ_i along with ID_G and λ_G of the gateway. In these schemes, each C_i is connected to its S_j via a gateway. Moreover, each S_j in these schemes incur excessive storage overhead as they need to store the security primitives of n clients and m gateways. As discussed earlier, λ_i is of 128 b. Thus, the cost incurred by S_j is n times higher than C_i for storing λ_i of n clients in [23]

TABLE V
STORAGE OVERHEAD COMPARISON

Schemes	Client Side	Server Side
<i>Li et. al [23]</i>	$(ID_i+\lambda_i)+(ID_G+\lambda_G)$	$n(ID_i+\lambda_i) + m(ID_G+\lambda_G)$
<i>Gupta et. al [6]</i>	-	-
<i>Gope et. al [7]</i>	-	-
<i>Chang et. al [9]</i>	$(ID_i+\lambda_i)+(ID_G+\lambda_G)$	$n(ID_i+\lambda_i) + m(ID_G+\lambda_G)$
<i>Proposed Scheme</i>	$ID_i+\lambda_i$	$n(ID_i+\lambda_i)$

TABLE VI
RESILIENCE AGAINST VARIOUS ATTACKS

Attacks	[23]	[6]	[7]	[9]	Proposed
<i>Replay</i>	Yes	Yes	Yes	Yes	Yes
<i>Eavesdropping</i>	Yes	Yes	Yes	Yes	Yes
<i>Forward & Backward Secrecy</i>	Yes	Yes	Yes	No	Yes
<i>Client Impersonation</i>	No	Yes	Yes	Yes	Yes
<i>Server Impersonation</i>	No	Yes	Yes	Yes	Yes
<i>Anonymity</i>	No	Yes	No	No	Yes
<i>DoS</i>	No	No	No	No	Yes

and it is m times higher than C_i for storing λ_G of m gateways. Similar to [23], the cost incurred by S_j is n times higher than C_i for storing λ_i of n clients, and it is m times higher than C_i for storing λ_G of m gateways.

In **Table VI**, the resilience of our scheme against various adversarial attacks is compared with the existing schemes. In our scheme, η_{client} and η_{server} are generated by a pseudorandom number R_i and appended to a timer T_i . This combination of T_i and R_i makes it extremely difficult for an adversary to replay messages. In our scheme, the use of one-time nonces η_{client} and η_{server} restrict the adversary from active eavesdropping. An A_k may compromise the μ_j ; however, the latter does not reveal any information about the previous or future sessions. This is mainly because μ_j is a one-time session key generated every time. Hence, forward and backward secrecy are maintained by our scheme. An A_k may intercept the exchanged handshake messages $\langle \lambda_i(ID_i)_{h()}, \gamma_{challenge}, \beta_{challenge}, \gamma_{response} \rangle$ and may generate different message patterns such as $\langle \lambda_k(ID_k)_{h()}, \gamma_{challenge}^k, \beta_{challenge}^k, \gamma_{response}^k \rangle$. The A_k may impersonate as C_i by transmitting $\lambda_k(ID_k)_{h()}$, and $\beta_{challenge}^k$ to S_j . Also, the same A_k impersonates as S_j by transmitting $\gamma_{challenge}^k$ and $\gamma_{response}^k$ to C_i . To impersonate as C_i or S_j , A_k would need λ_i . Because, A_k fabricates its own λ_k that does not exist either with C_i or S_j , i.e., $\lambda_k \neq \lambda_i$, hence it is unable to launch client or server impersonation attack. Moreover, A_k would need to fabricate η_k , μ_k , and ID_k as well to launch these attacks. These parameters are computationally inefficient to be calculated as each one would require 2^{128} attempts. In our scheme, the identities of C_i and S_j are masked in the messages $(\lambda_i(ID_i)_{h()}, \gamma_{challenge}, \beta_{challenge}, \text{ and } \gamma_{response})$. An A_k cannot interpret the identities of C_i

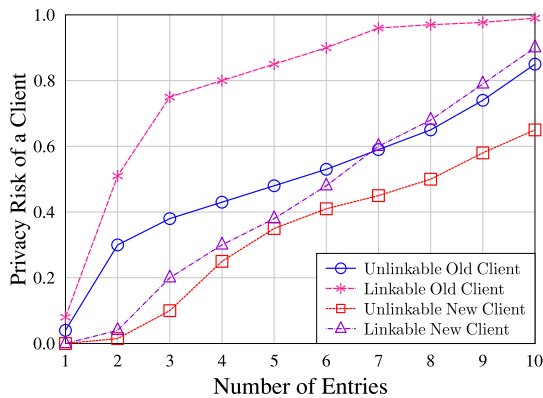


Fig. 4. Privacy risk prediction against number of entries.

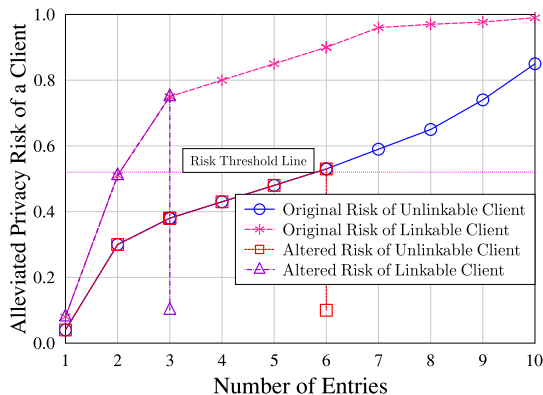


Fig. 5. Privacy risk alleviation against number of entries.

and S_j from the aforementioned messages as they are protected upon encryption by λ_i and μ_j . As a result, the anonymity of C_i and S_j is preserved. Moreover, our proposed scheme uses fresh nonces, i.e., η_{client} and η_{server} for every new session, and a new timer T_i as well. Hence, all sessions are nonlinkable and A_k is unable to trace any C_i and S_j from previous messages, thus providing untraceability feature. Finally, S_j restricts a C_i to only one connection at a given time. As a result, it is extremely difficult for an adversary to launch a DoS attack. In comparison to our scheme, all the existing schemes are susceptible to one or more such attacks and affect the privacy of C_i and S_j in one way or the other.

Our proposed scheme has used HMM to predict the privacy leakage of a client. In Fig. 4, we have shown the client's privacy risk against the number of entries. The privacy risk is associated with the number of visits, i.e., entries, a client makes to a hospital server. As evident from this figure, the privacy of linkable clients is higher than unlinkable clients, where a linkable client is the one whose personal identification can be extracted from entries and search results on a particular topic. For example, when a client searched for a specialist practitioner and read his/her profile or read about a particular disease etc. When a client visits the hospital server for the first time, his/her privacy risk is relatively low and increases with each entry to the hospital server. If the personal identification of this new client is linkable, the privacy risk is higher in comparison to unlinkable client.

Similarly, for old linkable clients, the privacy risk is highest and is moderate for unlinkable clients. The proposed scheme preserve the privacy of clients by predicting the privacy leakage using HMM. When the predicted privacy leakage crosses a specified threshold, the risk is altered, as shown in Fig. 5. The threshold is probabilistic and application-dependent that can be changed according to the application requirements. In this article, the threshold probability is 0.52, and once privacy leakage crosses it, the client's information is altered, and risk is alleviated, as shown in Fig. 5.

VI. CONCLUSION

In this article, we proposed a lightweight mutual authentication and key establishment scheme for IIoT wearable devices of I-CPS. The proposed scheme was based on client-server interaction model that used symmetric encryption. It was extremely lightweight and was suitable for large-scale I-CPS infrastructures. It was feasible for clients having limited resources and requires low computational, communication, and storage overhead while interacted with the servers for the exchange of session keys. After authentication, the privacy leakage of clients and their data was predicted using HMM. Upon privacy leakage detection, the data were altered through semantic similarity function with a loss in utility. The efficiency, correctness, and robustness of the security scheme were analyzed using BAN logic. The analysis showed that the proposed scheme was highly resilient against various adversarial attack. Moreover, it was efficient in terms of computation, communication, and storage overhead due to lightweight primitives, fewer number of exchanged messages and the absence of gateways, respectively. In the future, we aimed to use software-defined network for analyzed the exchanged data and the behavior of interacting entities of our scheme.

REFERENCES

- [1] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, "Next-generation big data analytics: State of the art, challenges, and future research topics," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1891–1899, Aug. 2017.
- [2] J. Xu, L. Wei, W. Wu, A. Wang, Y. Zhang, and F. Zhou, "Privacy-preserving data integrity verification by using lightweight streaming authenticated data structures for healthcare cyber-physical system," *Future Gener. Comput. Syst.*, vol. 108, pp. 1287–1296, 2018.
- [3] F. Al-Turjman and S. Alturjman, "Context-sensitive access in industrial internet of things (IIoT) healthcare applications," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2736–2744, Jun. 2018.
- [4] E. Luo, M. Z. A. Bhuiyan, G. Wang, M. A. Rahman, J. Wu, and M. Atiqzaman, "Privacyprotector: Privacy-protected patient data collection in IoT-based healthcare systems," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 163–168, Feb. 2018.
- [5] H. Tao, M. Z. A. Bhuiyan, A. N. Abdalla, M. M. Hassan, J. M. Zain, and T. Hayajneh, "Secured data collection with hardware-based ciphers for IoT-based healthcare," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 410–420, Feb. 2019.
- [6] A. Gupta, M. Tripathi, T. J. Shaikh, and A. Sharma, "A lightweight anonymous user authentication and key establishment scheme for wearable devices," *Comput. Netw.*, vol. 149, pp. 29–42, 2019.
- [7] P. Gope and T. Hwang, "A realistic lightweight anonymous authentication protocol for securing real-time application data access in wireless sensor networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7124–7132, Nov. 2016.
- [8] J. Srinivas, A. K. Das, N. Kumar, and J. Rodrigues, "Cloud centric authentication for wearable healthcare monitoring system," *IEEE Trans. Dependable Secure Comput.*, vol. 17, no. 5, pp. 942–956, Sep./Oct. 2020.

- [9] C.-C. Chang and H.-D. Le, "A provably secure, efficient, and flexible authentication scheme for ad hoc wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 357–366, Jan. 2016.
- [10] S. R. Moosavi *et al.*, "End-to-end security scheme for mobility enabled healthcare Internet of Things," *Future Gener. Comput. Syst.*, vol. 64, pp. 108–124, 2016.
- [11] J. Liu, Z. Zhang, X. Chen, and K. S. Kwak, "Certificateless remote anonymous authentication schemes for wireless body area networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 2, pp. 332–342, Feb. 2014.
- [12] A. K. Das, M. Wazid, N. Kumar, M. K. Khan, K.-K. R. Choo, and Y. Park, "Design of secure and lightweight authentication protocol for wearable devices environment," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1310–1322, Jul. 2018.
- [13] R. Canetti, O. Goldreich, and S. Halevi, "The random oracle methodology, revisited," *J. ACM*, vol. 51, no. 4, pp. 557–594, 2004.
- [14] M. Burrows, M. Abadi, and R. M. Needham, "A logic of authentication," *Proc. Roy. Soc. London A. Math. Phys. Sci.*, vol. 426, no. 1871, pp. 233–271, 1989.
- [15] K. Fritchman *et al.*, "Privacy-preserving scoring of tree ensembles: A novel framework for ai in healthcare," in *Proc. IEEE Int. Conf. Big Data*, 2018, pp. 2413–2422.
- [16] I. Bartoletti, "Ai in healthcare: Ethical and privacy challenges," in *Proc. Conf. Artif. Intell. Med. Eur.*, 2019, pp. 7–10.
- [17] T.-T. Kuo and L. Ohno-Machado, "Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks," Cornell Univ., 2018, pp. 1–13, *arXiv:1802.01746*.
- [18] S. Wang *et al.*, "A fast CP-ABE system for cyber-physical security and privacy in mobile healthcare network," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 4467–4477, Jul./Aug. 2020.
- [19] J. Duan, J. Hurd, G. Li, S. Owens, K. Slind, and J. Zhang, "Functional correctness proofs of encryption algorithms," in *Proc. Int. Conf. Log. Program. Artif. Intell. Reasoning*, 2005, pp. 519–533.
- [20] R. Masood, D. Vatsalan, M. Ikram, and M. A. Kaafar, "Incognito: A method for obfuscating web data," in *Proc. World Wide Web Conf.*, 2018, pp. 267–276.
- [21] L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo, and M. Palaniswami, "PPFA: Privacy preserving fog-enabled aggregation in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3733–3744, Aug. 2018.
- [22] S. Roy, S. Chatterjee, A. K. Das, S. Chattopadhyay, S. Kumari, and M. Jo, "Chaotic map-based anonymous user authentication scheme with user biometrics and fuzzy extractor for crowdsourcing Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2884–2895, Aug. 2018.
- [23] X. Li, M. H. Ibrahim, S. Kumari, A. K. Sangaiah, V. Gupta, and K.-K. R. Choo, "Anonymous mutual authentication and key agreement scheme for wearable sensors in wireless body area networks," *Comput. Netw.*, vol. 129, pp. 429–443, 2017.



Mian Ahmad Jan (Senior Member, IEEE) received the Ph.D. degree in computer systems from the University of Technology Sydney (UTS), Ultimo, NSW, Australia, in 2016.

He is currently an Assistant Professor with the Department of Computer Science, Abdul Wali Khan University Mardan, Mardan, Pakistan. He has been actively involved in ML, big data analytics, smart cities infrastructure and vehicular ad hoc networks. He has authored or coauthored the IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE INTERNET OF THINGS JOURNAL, IEEE JOURNAL OF SELECTED AREAS OF COMMUNICATIONS and ACM COMPUTING SURVEYS are few to mention. His research interests include energy-efficient and secured communication in wireless sensor networks and Internet of Things.

Dr. Ahmad Jan had been the recipient of various prestigious scholarships during his Ph.D. studies. He was the recipient of International Research Scholarship, UTS and Commonwealth Scientific Industrial Research Organization scholarships. He has been awarded the best Researcher awarded for the year 2014 with the University of Technology Sydney Australia. He has been a Guest Editor of numerous special issues in various prestigious journals such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATION, *Springer Neural Networks and Applications*, and *Elsevier Future Generation Computer Systems*, etc.

Dr. Ahmad Jan had been the recipient of various prestigious scholarships during his Ph.D. studies. He was the recipient of International Research Scholarship, UTS and Commonwealth Scientific Industrial Research Organization scholarships. He has been awarded the best Researcher awarded for the year 2014 with the University of Technology Sydney Australia. He has been a Guest Editor of numerous special issues in various prestigious journals such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATION, *Springer Neural Networks and Applications*, and *Elsevier Future Generation Computer Systems*, etc.



Fazlullah Khan (Senior Member IEEE) received the Ph.D. degree in computer science from AWKUM, in 2019.

He is currently an Assistant Professor with the Computer Science Department, Abdul Wali Khan University Mardan, Mardan, Pakistan. He has been involved in latest developments in the field of Internet of Vehicles security and privacy issues, software-defined networks, fog computing and big data analytics. He has authored his research work in top-notch journals and conferences.

His research has been published in the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE INTERNET OF THINGS, IEEE ACCESS, *Elsevier Computer Networks*, *Elsevier Future Generations Computer Systems*, *Elsevier Journal of Network and Computer Applications*, *Elsevier Computers and Electrical Engineering*, *Springer Mobile Networks and Applications*. His research interests include intelligent and robust protocol designs, security and privacy of wireless communication systems, Internet of Things, ML, and AI.

Dr. Khan had been the recipient of various prestigious scholarships during his Ph.D. studies and has been awarded the best Researcher awarded for the year 2017. He has served more than ten conferences in leadership capacities including General Chair, General Co-Chair, Program Co-Chair, Track Chair, Session Chair, and Technical Program Committee member, including the IEEE International Conference on Trust, Security and Privacy in Computing and Communications 2017, 2018, EuroCom, Global Conference on Consumer Electronics 2019, International Conference on Information Technology: New Generations 2018, Future5V 2017, CCODE-2017, IoT-BC2 2016. He has been an active reviewer for high-cited and highly ranked international journals, including IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, *Elsevier Computer Networks*, *Springer Mobile Networks and Applications* and *Wiley Concurrency and Computation: Practice and Experience*.



Rahim Khan (Member, IEEE) received the Ph.D. degrees in computer system engineering from the Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Swabi, Pakistan, in 2016.

He is currently an Assistant Professor with the Computer Science Department, Abdul Wali Khan University Mardan, Mardan. His research interests include the wireless sensor networks deployment and routing protocols, outliers detection, congestion control, decision support

system, vehicular ad-hoc networks, and similarity measures.



Spyridon Mastorakis (Member, IEEE) received the M.S. degree in computer science from the University of California, Los Angeles (UCLA), Los Angeles, CA, USA, in 2017, and a five-year diploma (equivalent to M.Eng.) in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2014, and the Ph.D. degree in computer science from UCLA in 2019.

He is currently an Assistant Professor in computer science with the University of Nebraska Omaha, Omaha, Nebraska. His research interests include network systems and protocols, internet architectures, IoT and edge computing, and security.



Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017.

He is currently an Associate Professor in computer science engineering with SCMS Group of Educational Institutions, Ernakulam, India. He has authored more than 45 research papers in peer reviewed and highly indexed International Journals and Conferences. His research interests include information science, scientometrics, digital library management, informatics of scientific databases, educational psychology, cyber psychology, hijacked and predatory journals, ad-hoc networks, wireless communication, opportunistic routing, wireless sensor networks, Internet of Things, fog computing and networking, underwater acoustic sensor networks, evaluation methods in education, online education tools, life skills training, training and development.



Paul Watters (Senior Member, IEEE) received the Ph.D. degree in cyber security from Macquarie University, Sydney, NSW, Australia, in 2000.

He is currently an Adjunct Professor of Cybersecurity with La Trobe University, and Honorary Professor with Macquarie University. He is a Chartered IT Professional.

Dr. Watters is a fellow of the British Computer Society and a Member of the Australian Psychological Society. He is an Academic Dean at Australasian Academies Polytechnic, an ASX-listed education provider. He is also Australia's leading trusted cybersecurity advisor, thought leader, and founder of Cyberstronomy Pty Ltd.



Mamoun Alazab (Senior Member, IEEE) received the Ph.D. degree in computer science from the School of Science, Information Technology and Engineering, Federation University of Australia, Ballarat, VIC, Australia, in 2012.

He is currently an Associate Professor with the College of Engineering, IT and Environment, Charles Darwin University, Casuarina, NT. He is a cyber security Researcher and Practitioner with industry and academic experience. He has more than 150 research papers in many international journals and conferences, such as the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, IEEE TRANSACTIONS ON BIG DATA, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computers and Security*, and *Future Generation Computing Systems*. His research interests include multidisciplinary that focuses on cyber security and digital forensics of computer systems with a focus on cybercrime detection and prevention.

Dr. Alazab delivered many invited and keynote speeches, 24 events in 2019 alone. He convened and chaired more than 50 conferences and workshops. He works closely with government and industry on many projects, including Northern Territory (NT) Department of Information and Corporate Services, IBM, Trend Micro, the Australian Federal Police, the Australian Communications and Media Authority, Westpac, United Nations Office on Drugs and Crime, and the Attorney General's Department. He is the Founding Chair of the IEEE NT Subsection.

Dr. Alazab delivered many invited and keynote speeches, 24 events in 2019 alone. He convened and chaired more than 50 conferences and workshops. He works closely with government and industry on many projects, including Northern Territory (NT) Department of Information and Corporate Services, IBM, Trend Micro, the Australian Federal Police, the Australian Communications and Media Authority, Westpac, United Nations Office on Drugs and Crime, and the Attorney General's Department. He is the Founding Chair of the IEEE NT Subsection.

I/Q Imbalance Aware Nonlinear Wireless-Powered Relaying of B5G Networks: Security and Reliability Analysis

Xingwang Li¹, Senior Member, IEEE, Mengyan Huang², Student Member, IEEE, Yuanwei Liu³, Senior Member, IEEE, Varun G Menon⁴, Senior Member, IEEE, Anand Paul⁵, Senior Member, IEEE, and Zhiguo Ding⁶, Fellow, IEEE

Abstract—Physical layer security is known as a promising paradigm to ensure secure performance for the future beyond 5 G (B5G) networks. In light of this fact, this paper elaborates on a tractable analysis framework to evaluate the reliability and the security of wireless-powered decode-and-forward (DF) multi-relay networks. More practical, the nonlinear energy harvesters, in-phase and quadrature-phase imbalance (IQI) and channel estimation errors (CEEs) are taken into account. To further enhance the secure performance, two relay selection strategies are presented: 1) suboptimal relay selection (SRS); 2) optimal relay selection (ORS). Specifically, exact analytical expressions for the outage probability (OP) and the intercept probability (IP) are derived in closed-form. For the IP, we consider that the eavesdropper can wiretap the signal from the source or the relay. In order to obtain more deep insights, we carry out the asymptotic analysis as well as the diversity orders for the OP in the high signal-to-noise ratio (SNR) regimes. Numerical results show that: 1) Although the mismatches of amplitude/phase of transmitter (TX)/receiver (RX) limit the OP performance, it can enhance IP performance; 2) Large number of relays yields better OP performance; 3) There are error floors for the OP due to the CEEs; 4) There is a trade-off for the OP and IP to obtain the balance between reliability and security.

Index Terms—B5G, channel estimation error, in-phase and quadrature-phase imbalance, nonlinear energy harvester, physical layer security.

Manuscript received May 12, 2020; revised August 20, 2020; accepted August 28, 2020. **Date of publication September 3, 2020;** date of current version December 9, 2021. This work was supported in part by the Henan Scientific and Technological Research Project under Grant 182102210307 and Grant 202102210122, in part by the Fundamental Research Funds for the Universities of Henan Province under Grant NSFRF180309, in part by the Outstanding Youth Science Foundation of Henan Polytechnic University under Grant J2019-4, in part by the Key Scientific Research Projects of Colleges and Universities in Henan Province under Grant 20A510007, and in part by the NSFC of China under Grant 61601414. Recommended for acceptance by Dr. Neeraj Kumar. (Corresponding author: Xingwang Li.)

Xingwang Li and Mengyan Huang are with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China (e-mail: lixingwangbupt@gmail.com; huangmengyan66@163.com).

Yuanwei Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk).

Varun G Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683 576, India (e-mail: varunmenon@ieee.org).

Anand Paul is with the School of Computer Science and Engineering, Kyungpook National University, 702701, South Korea (e-mail: paul.editor@gmail.com).

Zhiguo Ding is with the School of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, U.K. (e-mail: zhiguo.ding@manchester.ac.uk).

Digital Object Identifier 10.1109/TNSE.2020.3020950

2327-4697 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

I. INTRODUCTION

THE goals of the future fifth generation (5 G) and beyond (B5G) wireless networks can provide reliable communication among almost all aspects of life through the networks with higher data rate, lower latency and ubiquitous connectivity [1]. The security has been identified as a vital factor for wireless communication systems, which has triggered enormous interests from both academia and industry [2]–[4]. However, due to the broadcast characteristics of wireless propagation environments, it is a great challenge to ensure secure communication for the wireless networks without being eavesdropped by un-authorized receivers. The conventional methods are to use encryption algorithms, which impose extra computational overhead and system complexity [5]. In addition, with the rapid development of chip and computer technologies, conventional encryption technologies can not provide perfect security.

As an alternative way, physical layer security (PLS) has sparked a great deal of research interests [2]. The basic principle of PLS is to exploit the inherent randomness of fading channels to resist the information to be extracted by eavesdroppers [3].

Recently, there are a great of research works investigated the PLS under various fading channels, e.g. see [6]–[8] and the references therein. In [6], a secure transmit-beamforming of the multiple-input multiple-output (MIMO) systems over Rayleigh fading channels was designed, in which the maximal ratio combining (MRC) receivers were adopted to maximize the signal-to-noise ratio (SNR) at the main receiver.

Meanwhile, the secure performance of the classic wiretap model was investigated over the generalized Gamma fading channels and the analytical expressions of the strictly positive secrecy capacity (SPSC) probability and the lower bound for the secrecy outage probability (SOP) were derived [7]. On the other hand, considering $\kappa - \mu$ shadowed fading channels, the authors investigated the secrecy performance of classic Wyner's wiretap model and the approximate expressions on the lower bound for the SOP in the high SNR region and SPSC probability were derived with the aid of a moment matching method [8].

In real communication environments, it is difficult to have direct links between sources and destinations due to shadow fading and/or obstacle, so it is indispensable to use the relay to complete the communication [9]. In light of this fact, relaying

assisted transmission has been identified as one of the key technologies in the current and future wireless cooperative networks [10], [11]. The signals can be decoded and transferred from the source to the destination by using low cost and low power consumption relay nodes. In general, there are two basic relay protocols: i) amplify-and-forward (AF) [12]–[14], and ii) decode-and-forward (DF) [15], [16]. In [12], the outage probability (OP) performance is investigated based on the mobile device-to-device cooperative networks with incremental AF relaying and transmit antenna selection. The authors of [13] considered N -Nakagami- m fading channels and analyzed the average bit error probability (ABEP) of AF relaying networks. In [14], the authors investigated the performance of a multi-hop AF communication network over Nakagami-0.5 channels and the closed-form analytical approximate expressions for the OP and EC were obtained. To maximize confidentiality, the authors in [15] investigated the secure performance of multiple DF relay systems.

When deploying multiple relays in the systems, it will incur extra inter-relay (IR) interference and energy consumption. To avoid this problem, relay selection (RS) technology has been proposed [17], where some relay are selected according to some principle. Among the various RS schemes, optimal relay selection (ORS), suboptimal relay selection (SRS) and MRC are the most prevalent ones [18]–[20]. The pioneering work of the ORS scheme has been proposed by Bletsas according to selecting the relay with the largest instantaneous end-to-end SNR [21]. Based on the ORS, the authors in [18] investigated the symbol error rate (SER) of AF relay systems. To reduce the requirement of channel knowledge, the authors proposed a SRS scheme that the optimal relay is selected according to the link either source-relay or relay-destination [19]. Cognitive radio inspired cooperative relay systems was introduced, and the secure outage performance was studied over independent and non-identically distributed Nakagami- m fading channels. In [20], the authors compared the SOP of cognitive radio networks for ORS, SRS with MRC schemes over Nakagami- m fading channels.

Although the performance of wireless cooperative networks can be improved by appropriate relay protocols and RS schemes, the operation of wireless communication systems is constrained by power shortages of their wireless devices. This happens that in some cases the nodes are deployed in the remote or power limited areas [22]. Motivated by this fact,

some energy harvesting (EH) techniques have been proposed to prolong the life of the batteries of such wireless transmission devices [23], [24]. Among the various EH techniques, radio frequency (RF) enabled simultaneous wireless information and power transfer (SWIPT) attracts a lot of attentions because it can overcome the limitations of some other renewable energy resources such as solar energy, wind energy and magnetic induction that can only be used in some specific circumstances [25]. In addition, RF signals are ubiquitous in electromagnetic waves, and EH in RF is green, safe, controllable and reliable [26]. In general, there are usually two common protocols for SWIPT systems: i) time-switching (TS) and ii) power-splitting (PS) [27]–[30]. For PS, the authors proposed an artificial-noise(AN)-aided transmission scheme to facilitate the secure information transmission to information receivers

(IRs) and meet the EH requirement for energy receivers (ERs) [27]. The authors of [28] considered an underlay cognitive radio system based on the PS protocol SWIPT technique, and derived the expressions of upper and lower bounds of probability of SPSC.

In addition, most of the existing works assumed the linear energy harvester mode in the SWIPT systems. Recent works in [31]–[33] showed that the nonlinear mode is more practical, because the electronic devices used in the power conversion circuit are nonlinear. Furthermore, the linear mode of the energy harvester may cause serious resource allocation mismatch, resulting in a significant performance degradation. It is also a special case of the nonlinear mode when the saturation threshold of the energy harvester is infinite, i.e., unlimited battery capacity. Therefore, it is of high practical relevance to consider a nonlinear energy harvester.

Unfortunately, the aforementioned studies are based on the assumption of ideal hardware components and perfect channel state information (CSI), which is unrealistic in practical wireless communication systems. In practice, due to component mismatch and manufacturing non-idealities, these monolithic architectures inevitably have defects associated with the RF front-ends, thereby limiting the overall system performance [34]. A typical example is the in-phase and quadrature-phase imbalance (IQI), which refers to the mismatches of amplitude and phase between I and Q branches of the transceiver. This will result in incomplete image suppression [35]. Ideally, the I and Q branches of the mixer have an amplitude of 0 and a phase shift of 90° , providing an infinitely attenuated image band; however, in practice, the transceiver is susceptible to some analog front-end damage, and these damages introduce errors in the phase shift resulting in amplitude mismatch between the I and Q branches, thereby damaging the down-converted signal constellation, thereby increasing the corresponding error [36].

Motivated by the above practical concern, several research works have studied the systems secure performance in the presence of IQI [37]–[39]. Under the assumption of uncorrelation between channel of each subcarrier and its image, Ozdemir *et al.* in [37] derived an exact expression for the SINR of OFDM systems with IQI at transceivers. The authors analyzed the impact of joint IQI on the security and the reliability of cooperative NOMA for IoT Networks [38]. Considering backscatter communication, Li *et al.* in [39] derived analytical expressions for OP and the intercept probability (IP) of ambient backscatter NOMA systems under IQI.

On the other hand, imperfect CSI (ICSI) may be existed due to the presence of channel estimation errors (CEEs) and feedback delay. Therefore, it is of great practical significance to study the impact of ICSI and IQI on the security performance of cooperative networks.

A. Motivation and Contribution

Specifically, we investigate the reliability and security by deriving the analytical expressions for the OP and IP. Regarding to security, the direct transmission and cooperative transmission through relay are considered. In this study, we assume that the source and relay nodes of the networks are configured with

nonlinear energy harvesters to harvest energy from the nearby power beacon under different saturation thresholds. This is reasonable in some applications, such as internet-of-things (IoT), mesh networks and Ad Hoc networks, etc. The main contributions of this paper are summarized as follows:

- Considering IQI and CEEs, we consider three representative RS schemes, namely RRS, SRS and ORS. RRS is considered as a benchmark for the purpose of comparison. In SRS, the optimal relay is selected according to the channel conditions either the $S \rightarrow R_m$ or the $R_m \rightarrow D$. In ORS, the optimal relay is selected according to the link qualities both the $S \rightarrow R_m$ and the $R_m \rightarrow D$. The major difference between our work and [40] is that to study the effects of IQI caused by the mismatches of amplitude and phase between I and Q branch.
- Different from the most existing research works, this study adopts a more realistic nonlinear EH model due to the nonlinearity characteristic of the electronic devices [41], [42]. We have the assumption that nonlinear energy harvesters are equipped at the source and relays, which can harvest energy from the nearby power beacon by using TS protocol.
- For the reliability, we derived the exact analytical expressions for the OP of the proposed system for the three RS schemes. For the security, we consider two typical scenarios that direct transmission and cooperative transmission, the exact closed-form analytical expressions of the IP for the two scenarios are derived.¹
- To obtain more insights, we derived the asymptotic analytical expressions and diversity orders for the OP of the three RS schemes under non-ideal conditions. It reveals that there are error floors for the OP due to the non-zero CEEs, and the OP performance is limited by the IQI parameters.

B. Organization and Notations

The rest of the paper is organized as follows. In section II, we present a brief introduction of the considered system model. In section III, the reliability of the considered system for the three RS schemes is studied in terms of OP, while the security is analyzed through deriving the analytical expressions for the IP. In section IV, some numerical results are provided to verify the correctness of our analysis. Finally, we present a conclusion of this paper in Section V.

We use $|\cdot|$ to define absolute value. The notations $E\{\cdot\}$ and \triangleq denote the expectation and definition operations, respectively. $e \sim \mathcal{CN}(\mu, \sigma^2)$ defines a complex Gaussian distribution with a mean of μ and a variance of σ^2 . $\Pr\{\cdot\}$ represents the probability and $K_v(\cdot)$ denotes the v -th order modified Bessel function of the second kind. The probability density function (PDF) and cumulative distribution function (CDF) are expressed by $f_X(\cdot)$ and $F_X(\cdot)$, respectively. Finally, $\log_2(\cdot)$ is the logarithm.

¹ In some cases, the eavesdropper can simultaneously receive signals from both source and relays. Our work can be easily extended to these cases by combining the received signals from source and relays by using the selection combining or MRC.

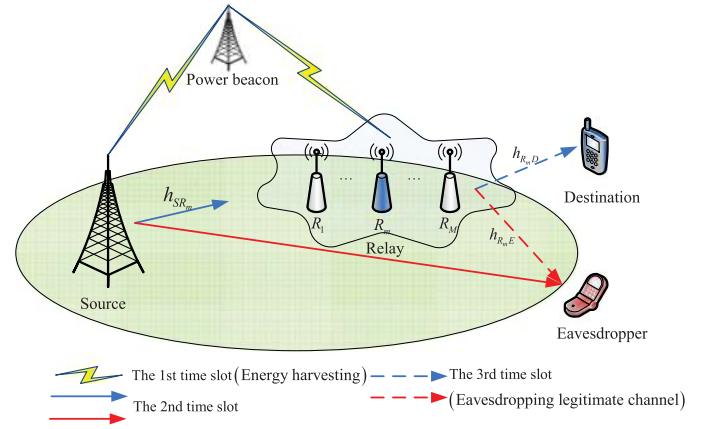


Fig. 1. System Model.

II. SYSTEM MODEL

We consider a DF multi-relay system as shown in Fig. 1, which deploys one power beacon B , one source S , M relays $R_m, m \in \{1, 2, \dots, M\}$, one destination D and one eavesdropper E . We assume that all nodes equipped with a single antenna and all nodes are operate in half-duplex (HD) mode. In order to improve the secure performance, the SRS and ORS schemes are designed to select the optimal relay among the M relays, while RRS scheme is presented as a benchmark. The source and all relay nodes are energy-constrained and can harvest energy from nearby B according to the TS protocol. It is considered that there is no direct link of $S \rightarrow D$ due to the blockage or heavy shadowing.

It is a great challenge to obtain perfect CSI in the communication process because of the CEEs, and the most common method is to estimate the channel using the training sequence. In this study, the linear minimum mean square error (LMMSE) is adopted here. Thus, the channel can be modeled as

$$h_j = \hat{h}_j + e_j, \quad (1)$$

where $\hat{h}_j, j \in \{SR_m, SE, R_mD, R_mE\}, (1 \leq m \leq M)$ is estimated channel of the real channel h_j , $e_j \sim \mathcal{CN}(0, \sigma_{e_j}^2)$ is the CEE, where $\sigma_{e_j}^2$ is the variance of estimation error, which is considered in two representative channel estimation models: 1) It is a non-negative fixed constant; 2) It is a function of the transmit average SNR and can be modeled as $\sigma_{e_j}^2 = \Omega_j(1 + \delta\rho_j\Omega_j)$, where δ is the channel estimation quality parameter that indicates the power consumption of the training pilot to obtain CSI; Ω_j and ρ_j are the variance of channel gain and transmit average SNR, respectively [43]. We assume that all communication links are subject to Rayleigh fading channels [44].

Typically, IQI is modeled as the phase and amplitude imbalance between transceiver I and Q signal paths. As depicts in [45], [46], the asymmetrical IQI model is considered, where I branch and Q branch are assumed to be ideal and errors, respectively. In this study, both transmitter (TX) and receiver (RX) are subject to IQI, in which case the transmitted baseband signals can be expressed as

$$x_{IQI} = \mu_{t/r}y_j + v_{t/r}y_j^*, \quad (2)$$

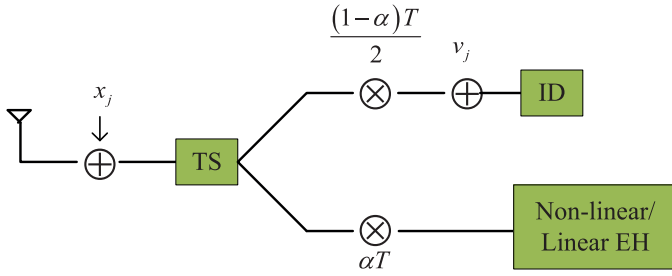


Fig. 2. Time-splitting structure.

where $y_j = \sqrt{P_{S/R}}x_j$ is the baseband signal that is transmitted under the conditions of non-ideal I/Q matching with $E\{|x_j|^2\} = 1$, x_j is the transmit signal of the TX; P_S and P_{R_m} are the transmit power at S and R_m , respectively; IQI coefficients are given by $\mu_t = \frac{1}{2}(1 + \xi_t \exp(j\phi_t))$, $v_t = \frac{1}{2}(1 - \xi_t \exp(-j\phi_t))$, $\mu_r = \frac{1}{2}(1 + \xi_r \exp(-j\phi_r))$, $v_r = \frac{1}{2}(1 - \xi_r \exp(j\phi_r))$, where ξ_{tr} and ϕ_{tr} denote the amplitude and phase mismatch at the TX and RX, respectively [47]. For ideal conditions, the parameters are set to $\xi_{tr} = 1$ and $\phi_{tr} = 0^\circ$ [48].

The entire data transmission is completed in three phases: 1) S and relays harvest energy from B ; 2) S transmits its own signals to R_m and E ; 3) R_m decodes and forwards the signals to D and E .

1) *The first phase:* In this phase, S and R_m are equipped with nonlinear harvesters can reap energy from B . As shown in Fig. 2, EH in the first duration of αT , the harvested energy at S and R_m can be formulated as

$$E_i = \varsigma_i P_B |h_{Bi}|^2 \alpha T, i \in \{S, R_m\} \quad (3)$$

where $\varsigma_{BS} \in (0, 1]$ and $\varsigma_{BR_m} \in (0, 1]$ are the energy converse coefficients of harvesters at S and R_m , respectively; P_B is the transmitted power at B ; h_{BS} is the channel coefficient between B and S ; h_{BR_m} is the channel coefficient between B and R_m ; α is the time allocation factor for EH, and T is the block transmission duration. The harvested energy E_i is used for information transmission in the second phase. The transmit power P_i can be expressed as follows in the case of the nonlinear energy harvester [41]

$$P_i = \begin{cases} \frac{2\alpha\varsigma_{Bi}P_B}{1-\alpha} |h_{Bi}|^2, & \text{if } P_B |h_{Bi}|^2 \leq \Gamma_i \\ \frac{2\alpha\varsigma_{Bi}}{1-\alpha} \Gamma_i, & \text{if } P_B |h_{Bi}|^2 > \Gamma_i \end{cases}, i \in \{S, R_m\} \quad (4)$$

where Γ_{BS} and Γ_{BR_m} are the saturation thresholds of the harvester at S and R_m , respectively.

2) *The second phase:* In this phase, information reception at the relays in the second duration of $\frac{(1-\alpha)T}{2}$, S respectively sends the signals x_{SR_m} and x_{SE} to R_m and E with $E\{|x_{SR_m}|^2\} = E\{|x_{SE}|^2\} = 1$. Considering IQI and CEEs, the received signals at R_m and E can be written as (5),

$$y_j = \mu_{r_j} \left[(\hat{h}_j + e_j) \left(\mu_{t_j} \sqrt{P_{S/R_m}} x_j + v_{t_j} \sqrt{P_{S/R_m}} x_j^* \right) + n_j \right] + v_{r_j} \left[(\hat{h}_j + e_j) \left(\mu_{t_j} \sqrt{P_{S/R_m}} x_j + v_{t_j} \sqrt{P_{S/R_m}} x_j^* \right) + n_j \right]^*, \quad (5)$$

where \hat{h}_{SR_m} and \hat{h}_{SE} are the estimated channel coefficients from transmitter to receiver; $n_{SR_m} \sim \mathcal{CN}(0, N_{SR_m})$ and $n_{SE} \sim \mathcal{CN}(0, N_{SE})$ are the complex additive white Gaussian noise (AWGN).

3) *The third phase:* In the third phase, information reception at D/E in the final duration of $\frac{(1-\alpha)T}{2}$, R_m respectively sends the signals x_{R_mD} , x_{R_mE} to D and E with $E\{|x_{R_mD}|^2\} = E\{|x_{R_mE}|^2\} = 1$. Similarly, the received signals at D and E can be expressed as (5).²

Hence, the received signal-to-interference-plus-noise ratio (SINRs) at R_m , D and E can be expressed in a unified form as

$$\gamma_j = \frac{|\hat{h}_j|^2 \rho_j p_j}{\sigma_{e_j}^2 \rho_j p_j + |\hat{h}_j|^2 \rho_j q_j + \sigma_{e_j}^2 \rho_j q_j + g_j}, \quad (6)$$

where $j \in \{SR_m, SE, R_mD, R_mE\}$, $\rho_j = P_{S/R}/N_j$, $p_j = |\mu_{t_j} \mu_{r_j} + v_{t_j}^* v_{r_j}|^2$, $q_j = |\mu_{r_j} v_{t_j} + \mu_{t_j}^* v_{r_j}|^2$ and $g_j = |\mu_{r_j} + v_{r_j}|^2$.

According to the Shannon's theorem, the channel capacity can be expressed as follows

$$C_j = \frac{1-\alpha}{2} \log_2(1 + \gamma_j), \quad (7)$$

where the factor $\frac{1-\alpha}{2}$ means the data transmission is accomplished in equal two phases.

With DF protocol, the effective end-to-end capacity from S to D can be expressed as

$$C_R = \min(C_{SR_m}, C_{R_mD}). \quad (8)$$

III. PERFORMANCE ANALYSIS

This section analyzes the reliability and security of considered system in the presence of nonlinear energy harvester, IQI and ICSI. The closed-form expressions for the OP under the RRS, SRS, ORS schemes and IP under direct transmission and relay transmission strategies are derived.³ [39]. Moreover, the asymptotic behaviors for the OP are examined, as well as the diversity orders.

A. Outage Probability Analysis

In the following, the expressions for the OP are presented according the three RS strategies considered IQI, ICSI and nonlinear energy harvesters. The OP is defined as the probability that effective channel capacity is below the threshold R_{th} , which can be expressed as

$$P_{out} \triangleq \Pr\{C_R < R_{th}\}. \quad (9)$$

1) *Random Relay Selection:* For RRS strategy, the link between S and arbitrary one of the relay R_m is selected, and the effective rate can be obtained as

² Note that $j \in \{SR_m, SE, R_mD, R_mE\}$, and P_S and P_{R_m} are the power from S and R_m , respectively.

³ The reliability and security are another metrics to characterize the PLS of wireless communication systems without using any secrecy coding, which are formulated by the OP and the IP

$$P_{out}^{RRS} = 1 - \left(\frac{\lambda_{SR_m}}{C_1} e^{-\frac{\lambda_{SR_m} C_2}{C_1}} \left[\sqrt{\frac{\beta_1}{\gamma_1}} K_1(\sqrt{\beta_1 \gamma_1}) - \frac{\pi \Lambda_1}{2 Y_1} \sum_{l=0}^{Y_1} e^{-\frac{2 \lambda_{BS} g_{SR_m} N_{SR_m} \varepsilon - \lambda_{SR_m} \Lambda_1 (\delta_l + 1)}{\Lambda_1 (\delta_l + 1)}} \frac{\lambda_{SR_m} \Lambda_1 (\delta_l + 1)}{2 C_1} \sqrt{1 - \delta_l^2} \right] + e^{\lambda_{SR_m} E_1} (e^{-\lambda_{SR_m} \Theta_1} - e^{-\lambda_{SR_m} T_1}) \right) \times \left(\frac{\lambda_{R_m D}}{C_3} e^{-\frac{\lambda_{R_m D} C_4}{C_3}} \left[\sqrt{\frac{\beta_2}{\gamma_2}} K_1(\sqrt{\beta_2 \gamma_2}) - \frac{\pi \Lambda_2}{2 Y_2} \sum_{l=0}^{Y_2} e^{-\frac{2 \lambda_{BR_m} g_{R_m D} N_{R_m D} \varepsilon - \lambda_{R_m D} \Lambda_2 (\delta_l + 1)}{\Lambda_2 (\delta_l + 1)}} \frac{\lambda_{R_m D} \Lambda_2 (\delta_l + 1)}{2 C_3} \sqrt{1 - \delta_l^2} \right] + e^{-\lambda_{R_m D} E_2} (e^{-\lambda_{R_m D} \Theta_2} - e^{-\lambda_{R_m D} T_3}) \right) \quad (11)$$

$$P_{out}^{SRS} = 1 + \left(\frac{\Xi}{C_5} e^{-\frac{\lambda_{SR_m} (s+1) C_6}{C_5}} \left[\sqrt{\frac{\beta_3}{\gamma_3}} K_1(\sqrt{\beta_3 \gamma_3}) - \frac{\pi \Lambda_3}{2 Y_3} \sum_{l=0}^{Y_3} e^{-\frac{2 \lambda_{BS} g_{SR_m} N_{SR_m} \varepsilon - \lambda_{SR_m} (s+1) \Lambda_3 (\delta_l + 1)}{\Lambda_3 (\delta_l + 1)}} \frac{\lambda_{SR_m} (s+1) \Lambda_3 (\delta_l + 1)}{2 C_5} \sqrt{1 - \delta_l^2} \right] - \frac{\Xi}{\lambda_{SR_m} (s+1)} e^{-\lambda_{SR_m} E_1 - \lambda_{SR_m} (s+1) T_5} - \left[1 - (1 - e^{-\lambda_{SR_m} \Theta_3})^M \right] e^{-\lambda_{SR_m} E_1} \right) \times \left(e^{-\lambda_{R_m D} E_2} (e^{-\lambda_{R_m D} \Theta_4} - e^{-\lambda_{R_m D} T_7}) + \frac{\lambda_{R_m D}}{C_7} e^{-\frac{\lambda_{R_m D} C_8}{C_7}} \left[\sqrt{\frac{\beta_4}{\gamma_4}} K_1(\sqrt{\beta_4 \gamma_4}) - \frac{\pi \Lambda_4}{2 Y_4} \sum_{l=0}^{Y_4} e^{-\frac{2 \lambda_{BR_m} g_{R_m D} N_{R_m D} \varepsilon - \lambda_{R_m D} \Lambda_4 (\delta_l + 1)}{\Lambda_4 (\delta_l + 1)}} \frac{\lambda_{R_m D} \Lambda_4 (\delta_l + 1)}{2 C_7} \sqrt{1 - \delta_l^2} \right] \right) \quad (19)$$

$$C_{R_m} = \min(C_{SR_m}, C_{R_m D}). \quad (10)$$

Based on the above discussion, we can obtain the analytical expression for the OP of the RRS strategy in the following theorem.

Theorem 1: The analytical expression for the OP of RRS strategy is provided in (11) as shown at the top of the page.

$$\text{where } A_1 = \frac{2\alpha\zeta_1}{1-\alpha}, E_1 = \frac{\Gamma_1}{P_B}, C_1 = A_1 P_B (p_{SR_m} - q_{SR_m} \varepsilon), C_2 = \sigma_{e_{SR_m}}^2 A_1 P_B \varepsilon (p_{SR_m} + q_{SR_m}), T_1 = \frac{g_{SR_m} N_{SR_m} \varepsilon}{C_1 E_1} + \frac{C_2}{C_1}, \beta_1 = 4\lambda_{BS} g_{SR_m} N_{SR_m} \varepsilon, \gamma_1 = \frac{\lambda_{SR_m}}{C_1}, \Lambda_1 = C_1 T_1 - C_2, \delta_{l_1} = \cos\left[\frac{(2l_1-1)\pi}{2Y_1}\right], \Theta_1 = \frac{\varepsilon \sigma_{e_{SR_m}}^2 A_1 \Gamma_1 (p_{SR_m} + q_{SR_m}) + \varepsilon g_{SR_m} N_{SR_m}}{A_1 \Gamma_1 (p_{SR_m} - \varepsilon q_{SR_m})}, A_2 = \frac{2\alpha\zeta_2}{1-\alpha}, E_2 = \frac{\Gamma_2}{P_B}, C_3 = A_2 P_B (p_{R_m D} - q_{R_m D} \varepsilon), C_4 = \sigma_{e_{R_m D}}^2 A_2 P_B \varepsilon (p_{R_m D} + q_{R_m D}), T_3 = \frac{g_{R_m D} N_{R_m D} \varepsilon}{C_3 E_2} + \frac{C_4}{C_3}, \beta_2 = 4\lambda_{BR_m} g_{R_m D} N_{R_m D} \varepsilon, \gamma_2 = \frac{\lambda_{R_m D}}{C_3}, \Lambda_2 = C_3 T_3 - C_4, \delta_{l_2} = \cos\left[\frac{(2l_2-1)\pi}{2Y_2}\right] \text{ and } \Theta_2 = \frac{\varepsilon \sigma_{e_{R_m D}}^2 A_2 \Gamma_2 (p_{R_m D} + q_{R_m D}) + \varepsilon g_{R_m D} N_{R_m D}}{A_2 \Gamma_2 (p_{R_m D} - \varepsilon q_{R_m D})}.$$

For $\varepsilon < 1 \max\{\frac{p_{SR_m}}{q_{SR_m}}, \frac{p_{R_m D}}{q_{R_m D}}\}$, otherwise the OP expressions are equal to 1.

Proof: See Appendix A. ■

To get deeper insights, the asymptotic behavior of non-ideal conditions ($\sigma_{e_{SR_m}}^2 = \sigma_{e_{R_m D}}^2 = t$) is investigated at high SNRs in the following corollary.⁴

Corollary 1: The asymptotic expression of OP for RRS strategy under non-ideal conditions ($\sigma_{e_{SR_m}}^2 = \sigma_{e_{R_m D}}^2 = t$) is given by

$$P_{out}^{RRS, \infty} = 1 - e^{-\lambda_{SR_m} H_1 - \lambda_{R_m D} H_2}, \quad (12)$$

where $H_1 = \varepsilon \sigma_{e_{SR_m}}^2 (p_{SR_m} + q_{SR_m}) / p_{SR_m} - \varepsilon q_{SR_m}$ and $H_2 = \varepsilon \sigma_{e_{R_m D}}^2 (p_{R_m D} + q_{R_m D}) / (p_{R_m D} - \varepsilon q_{R_m D})$.

Proof: Based on (7), the asymptotic channel capacities of $S \rightarrow R_m$ and $R_m \rightarrow D$ can be written in a unified form as

$$C_t^{\infty, ni} = \frac{1-\alpha}{2} \log_2 \left(1 + \frac{|\hat{h}_t|^2 p_t}{\sigma_{e_t}^2 p_t + |\hat{h}_t|^2 q_t + \sigma_{e_t}^2 q_t} \right), t \in \{SR_m, R_m D\} \quad (13)$$

⁴ Through this paper, we have $\xi_{tr} = 1$, $\phi_{tr} = 0^\circ$ and $\sigma_e^2 = 0$ for non-ideal conditions, while $\xi_{tr} \neq 1$, $\phi_{tr} \neq 0^\circ$ and $\sigma_e^2 = 0, \sigma_e^2 > 0$ for ideal conditions, respectively.

According to the definition of OP, the following expression can be obtained as

$$P_{out}^{\infty, ni} = \Pr\left\{ \min\left(C_{SR_m}^{\infty, ni}, C_{R_m D}^{\infty, ni}\right) < R_{th} \right\} = 1 - \Pr\left\{ C_{SR_m}^{\infty, ni} > R_{th} \right\} \Pr\left\{ C_{R_m D}^{\infty, ni} > R_{th} \right\}. \quad (14)$$

Utilizing the similar methodology of Appendix A, (12) can be derived. ■

Furthermore, the diversity order is investigated, which can be defined as [34]:

$$d = - \lim_{\rho_j \rightarrow \infty} \frac{\log(P_{out}^{\infty})}{\log \rho_j}, \quad (15)$$

where ρ_j is the average SNR and P_{out}^{∞} is the asymptotic OP.

Corollary 2: The diversity order of OP for RRS scheme in the presence of non-ideal conditions ($\sigma_{e_{SR_m}}^2 = \sigma_{e_{R_m D}}^2 = t$) can be obtained as follows:

$$d_{RRS}^{ni}(\rho_{SR_m}, \rho_{R_m D}) = 0. \quad (16)$$

Proof: Follows trivially by using (15) and the definition of derivative. ■

Remark 1: From **Theorem 1**, **Corollary 1** and **Corollary 2**, the following observations can be obtained as: 1) When M increases gradually, it can be seen that (11) and (12) are independent of M , so the RRS scheme will not change with the increase or decrease of the number of antennas; 2) At high average SNR, $P_{out}^{RRS, \infty}$ is a fixed non-zero constant, which results in 0 diversity order. This means that the diversity order can not be improved by increasing the number of relays.

2) *Suboptimal Relay Selection:* For SRS strategy, the optimal relay is selected according to maximizing the capacity of the link $S \rightarrow R_m$, which can be expressed as

$$a = \arg \max_{m=1,2,\dots,M} C_{SR_m}, \quad (17)$$

$$C_{R_a} = \min(C_{SR_a}, C_{R_a D}). \quad (18)$$

Based on (9) and (17), we have the following Theorem 2.

$$P_{out}^{ORS} = \prod_{m=1}^M \left\{ 1 - \left(\frac{\lambda_{SR_m}}{C_1} e^{-\frac{\lambda_{SR_m} C_2}{C_1}} \left[\sqrt{\frac{\beta_1}{\gamma_1}} K_1(\sqrt{\beta_1 \gamma_1}) - \frac{\pi \Lambda_1}{2 Y_1} \sum_{l_1=0}^{Y_1} e^{-\frac{2 \lambda_{BS} g_{SR_m} N_{SR_m} \varepsilon}{\Lambda_1 (\delta_{l_1+1})} - \frac{\lambda_{SR_m} \Lambda_1 (\delta_{l_1+1})}{2 C_1}} \sqrt{1 - \delta_{l_1}^2} \right] + e^{\lambda_{BS} E_1} (e^{-\lambda_{SR_m} \Theta_1} - e^{-\lambda_{SR_m} T_1}) \right) \right. \\ \left. \times \left(\frac{\lambda_{R_m D}}{C_3} e^{-\frac{\lambda_{R_m D} C_4}{C_3}} \left[\sqrt{\frac{\beta_2}{\gamma_2}} K_1(\sqrt{\beta_2 \gamma_2}) - \frac{\pi \Lambda_2}{2 Y_2} \sum_{l_2=0}^{Y_2} e^{-\frac{2 \lambda_{BR_m} g_{R_m D} N_{R_m D} \varepsilon}{\Lambda_2 (\delta_{l_2+1})} - \frac{\lambda_{R_m D} \Lambda_2 (\delta_{l_2+1})}{2 C_3}} \sqrt{1 - \delta_{l_2}^2} \right] + e^{-\lambda_{BR_m} E_2} (e^{-\lambda_{R_m D} \Theta_2} - e^{-\lambda_{R_m D} T_3}) \right) \right\} \quad (24)$$

Theorem 2: The analytical expression of OP is provided for SRS strategy in (19) as shown at the top of the previous page, where $\varepsilon = -M \lambda_{SR_a} \sum_{s=0}^{M-1} \binom{M-1}{s} (-1)^s$, $C_5 = A_1 P_B (p_{SR_a} - q_{SR_a} \varepsilon)$, $C_6 = \sigma_{e_{SR_a}}^2 A_1 P_B \varepsilon (p_{SR_a} + q_{SR_a})$, $T_5 = \frac{g_{SR_a} N_{SR_a} \varepsilon}{C_5 E_1} + \frac{C_6}{C_5}$, $\beta_3 = 4 \lambda_{BS} g_{SR_a} N_{SR_a} \varepsilon$, $\gamma_3 = \frac{\lambda_{SR_a} (s+1)}{C_5}$, $\Lambda_3 = C_5 T_5 - C_6$, $\delta_{l_3} = \cos \left[\frac{(2l_3-1)\pi}{2Y_3} \right]$, $C_7 = A_2 P_B (p_{R_a D} - q_{R_a D} \varepsilon)$, $C_8 = \sigma_{e_{R_a D}}^2 A_2 P_B \varepsilon (p_{R_a D} + q_{R_a D})$, $T_7 = \frac{g_{R_a D} N_{R_a D} \varepsilon}{C_7 E_2} + \frac{C_8}{C_7}$, $\beta_4 = 4 \lambda_{BR_a} g_{R_a D} N_{R_a D} \varepsilon$, $\gamma_4 = \frac{\lambda_{R_a D}}{C_7}$, $\Lambda_4 = C_7 T_7 - C_8$ and $\delta_{l_4} = \cos \left[\frac{(2l_4-1)\pi}{2Y_4} \right]$.

Proof: See Appendix B. ■

Similarly, the asymptotic behavior of non-ideal conditions is studied of OP for SRS strategy in the high SNR regime.

Corollary 3: The asymptotic expression for the OP of SRS strategy under non-ideal conditions ($\sigma_{e_{SR_a}}^2 = \sigma_{e_{R_a D}}^2 = t$) is given by

$$P_{out}^{SRS, \infty} = 1 - \left(1 - (1 - e^{-\lambda_{SR_a} H_3})^M \right) e^{-\lambda_{R_a D} H_4}, \quad (20)$$

where $H_3 = \varepsilon \sigma_{e_{SR_a}}^2 (p_{SR_a} + q_{SR_a}) / (p_{SR_a} - \varepsilon q_{SR_a})$ and $H_4 = \varepsilon \sigma_{e_{R_a D}}^2 (p_{R_a D} + q_{R_a D}) / (p_{R_a D} - \varepsilon q_{R_a D})$.

Then, the diversity order of OP for SRS strategy under non-ideal conditions ($\sigma_{e_{SR_a}}^2 = \sigma_{e_{R_a D}}^2 = t$) is presented in the following corollary.

Corollary 4: The diversity order of OP for SRS scheme in the presence of non-ideal conditions ($\sigma_{e_{SR_m}}^2 = \sigma_{e_{R_m D}}^2 = t$) is given by:

$$d_{SRS}^{mi}(\rho_{SR_a}, \rho_{R_a D}) = 0. \quad (21)$$

Remark 2: From **Theorem 2**, **Corollary 3** and **Corollary 4**, we can obtain the following conclusion as: 1) when the number of relay increases, it can be concluded from formulas (19) and (20) that the system's outage performance becomes better under the SRS strategy; 2) From expression (19), it can be obtained that when M is fixed and the transmit power at B is in a high state, the OP will cause an error floor; 3) From (21), we can observe that the diversity order of the considered system is zero due to the fixed constant for the OP in the high SNR regime.

3) **Optimal Relay Selection:** For ORS strategy, the optimal relay is selected according to maximize the capacity of the links both $S \rightarrow R_m$ and $R_m \rightarrow D$

$$m^* = \arg \max_{1 \leq m \leq M} \min \{ C_{SR_m}, C_{R_m D} \}, \quad (22)$$

$$C_{R_m^*} = \max_{1 \leq m \leq M} C_{R_m}. \quad (23)$$

According to (9) and (23), Theorem 3 can be obtained as following.

Theorem 3: The analytical expression of the OP is provided for the ORS strategy in (24) as shown at the top of the page.

Proof: See Appendix C. ■

Next, the asymptotic behavior for the OP of ORS strategy in the presence of non-ideal conditions is studied.

Corollary 5: The asymptotic expression of OP for the ORS strategy under non-ideal conditions ($\sigma_{e_{SR_a}}^2 = \sigma_{e_{R_a D}}^2 = t$) is given by

$$P_{out}^{ORS, \infty} = \prod_{i=1}^M (1 - e^{-\lambda_{SR_m} H_1 - \lambda_{R_m D} H_2}). \quad (25)$$

Corollary 6: The diversity order of OP for ORS strategy under non-ideal conditions ($\sigma_{e_{SR_m}}^2 = \sigma_{e_{R_m D}}^2 = t$) is following:

$$d_{ORS}^{mi}(\rho_{SR_m}, \rho_{R_m D}) = 0. \quad (26)$$

Remark 3: From **Theorem 3**, **Corollary 5** and **Corollary 6**, we can get the following points as: 1) When M increases, P_{out}^{ORS} and $P_{out}^{ORS, \infty}$ will become smaller because, which means that the system's outage performance becomes better under the ORS strategy; 2) As P_B goes to infinity, the OP of the considered system under non-ideal conditions has an error floor; 3) We can also observe that the diversity order is 0, which means that the slope of the outage probability is 0.

B. Intercept Probability Analysis

IP is an important performance metric of wireless communication systems, the secrecy performance of the multi-relay networks with IQI is studied in terms of IP considering two scenarios of direct transmission and transmission via relay. The definition of IP is the probability that the channel capacity between $S \rightarrow E$ or $R_m \rightarrow E$ is greater than the threshold R_{th} , which can be formulated as

$$P_{int}^{\text{direct/relay}} \triangleq \Pr \{ C_{SE/R_c E} > R_{th} \}, \quad (27)$$

where R_c is the selected relay, C_{SE} and $C_{R_c E}$ are the intercept capacities of $S \rightarrow E$ and $R_c \rightarrow E$, respectively.

1) **Direct Transmission:** Under the condition of direct transmission, based on (6) and the definition of (27), the closed-form analytical expression of IP under the condition of direct transmission can be obtained as Theorem 4.

Theorem 4: The analytical expression of IP under the condition of direct transmission is provided in (28) as shown at the bottom of the next page, where $C_9 = A_1 P_B (p_{SE} - q_{SE} \varepsilon)$,

TABLE I
 PARAMETERS FOR NUMERICAL RESULTS

Monte Carlo simulations repeated	10^7 iterations
Distance between nodes	$d_{R_m D} = d_{R_m E} = 1.5m, d_{SE} = 2m$
Shadow fading parameter	$\beta = 3$
Time allocation factor	$\alpha = 0.5$
Noise power	$N_{SR_m} = N_{SE} = N_{R_m D} = N_{R_m E} = 1$
Intercept capacity threshold	$R_{th} = 0.05$
Amplitude at TX and RX	$\xi_t = \xi_r = \{1, 1.1\}$,
Phase at TX and RX	$\phi_t = \phi_r = \{0^\circ, 5^\circ\}$
Variance of CEEs	$\varsigma_1 = \varsigma_2 = 0.5$
Energy converse coefficient at source and relay	$\sigma_e^2 = \{0, 0.05\}$

$C_{10} = \sigma_{e_{SE}}^2 A_1 P_B \varepsilon (p_{SE} + q_{SE})$, $T_9 = \varepsilon g_{SE} N_{SE} / C_9 E_1 + C_{10} / C_9$,
 $\beta_5 = \lambda_{BS} g_{SE} N_{SE} \varepsilon$, $\gamma_5 = \frac{\lambda_{SE}}{C_9}$, $\Lambda_5 = C_9 T_9 - C_{10}$, $\delta_{l_5} = \cos$
 $[(2l_5 - 1)\pi / 2Y_5]$ and $\Theta_5 = \varepsilon \sigma_{e_{SE}}^2 A_1 \Gamma_1(p_{SE} + q_{SE}) + \varepsilon g_{SE} N_{SE} /$
 $A_1 \Gamma_1(p_{SE} - \varepsilon q_{SE})$.

Proof: See Appendix D. ■

2) *Transmission via Relay:* We then studied the security of the considered system by utilizing relay to transmit information in the following theorem.

Theorem 5: The analytical expression of IP under the transmission via relay condition is provided in (29)

$$P_{int}^{\text{relay}} = e^{-\lambda_{BR_c} E_2} (e^{-\lambda_{R_c E} \Theta_6} - e^{-\lambda_{R_c E} T_{11}}) + \frac{\lambda_{R_c E}}{C_{11}} e^{-\frac{\lambda_{R_c E} C_{12}}{C_{11}}} \left[\sqrt{\frac{\beta_6}{\gamma_6}} K_1 \sqrt{\beta_6 \gamma_6} - \frac{\pi \Lambda_6}{2Y_6} \sum_{l_6=0}^{Y_6} e^{-\frac{\beta_6}{2\Lambda_6 (\delta_{l_6+1})} \frac{\gamma_6 \Lambda_6 (\delta_{l_6+1})}{2}} \sqrt{1 - \delta_{l_6}^2} \right] \quad (29)$$

where $C_{11} = A_2 P_B (p_{R_c E} - q_{R_c E} \varepsilon)$, $C_{12} = \sigma_{e_{R_c E}}^2 A_2 P_B \varepsilon$
 $(p_{R_c E} + q_{R_c E})$, $T_{11} = g_{R_c E} N_{R_c E} \varepsilon / C_{11} E_2 + C_{12} / C_{11}$, $\beta_6 =$
 $4\lambda_{BR_c} g_{R_c E} N_{R_c E} \varepsilon$, $\gamma_6 = \lambda_{R_c E} / C_{11}$, $\Lambda_6 = C_{11} T_{11} - C_{12}$, $\delta_{l_6} =$
 $\cos [(2l_6 - 1)\pi / 2Y_6]$ and $\Theta_6 = \varepsilon \sigma_{e_{R_c E}}^2 A_2 \Gamma_2(p_{R_c E} + q_{R_c E})$
 $+ \varepsilon g_{R_c E} N_{R_c E} / A_2 \Gamma_2(p_{R_c E} - \varepsilon q_{R_c E})$.

Proof: See Appendix E. ■

IV. NUMERICAL RESULTS

In this section, some numerical results are provided to validate the correctness of the obtained results in the above section. The results are then verified using Monte Carlo simulations with 10^7 iterations. Unless otherwise specified, we set the parameters as in Table I.

A. Reliability Analysis

Fig. 3 plots the OP versus the transmit power P_B for different RS strategies. For the purpose of comparison, the curves of ideal conditions are provided. We set $M = 2$ [40]. These simulation results perfectly verify the derived closed-form analytical expressions of (11), (19) and (24) and asymptotic expressions of (12), (20) and (25), as well as (16), (21) and (26). We can also see from the simulation results that: 1) OP

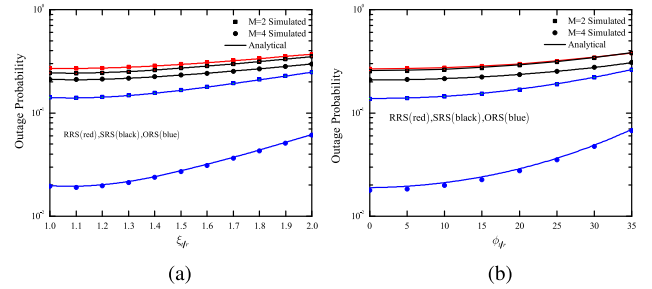


Fig. 3. Influence of IQI: (a) OP versus TX/RX amplitude; (b) OP versus phase mismatch.

under the non-ideal conditions is greater than that of the ideal conditions due to the IQI and CEEs; 2) The outage performance under RRS strategy is worse than SRS and ORS strategies, and ORS scheme has the best outage performance; 3) There are error floors of the OP for the three RS schemes in the high regions due to CEEs, which means that the system OP performance can not always be improved by increasing the transmit power.

Fig. 4(a) illustrates the OP of TX/RX amplitude ξ_{tr} for different number of relays ($M = \{2, 4\}$) under two RS strategies [49]. We set $P_B = 20$ dB. These results indicate that the OP for SRS is higher than that of ORS for arbitrary number of relay ($M > 1$). This gap of OP between the two schemes becomes large as the number of relays increases. Also, we can see that the outage performance of the considered system is proportional to the M . Finally, the OP of the system increases gradually with the increase of TX/RX amplitude, which means that the ξ_{tr} has negative effects on the system performance. Fig. 4(b) plots the OP versus phase mismatch ϕ_{tr} for different number of relays under two RS strategies. As in Fig. 4(a), we set $P_B = 20$ dB. These simulation results verify that with the increase of ϕ_{tr} , the outage performance of the system gradually becomes worse. Furthermore, the effects of the number of relays on the system performance in Fig. 3 and Fig. 4 are further verified. From Fig. 4(a) and Fig. 4(b), we can observe that the parameters of amplitude and phase mismatches have the similar effects on the outage performance.

$$P_{int}^{\text{direct}} = -e^{-\lambda_{BS} E_1 - \lambda_{SE} T_9} + \frac{\lambda_{SE}}{C_9} e^{-\frac{\lambda_{SE} C_{10}}{C_9}} \left[2 \sqrt{\frac{\beta_5}{\gamma_5}} K_1 \left(2 \sqrt{\beta_5 \gamma_5} \right) - \frac{\pi \Lambda_5}{2Y_5} \sum_{l_5=0}^{Y_5} e^{-\frac{\gamma_5 \Lambda_5 (\delta_{l_5+1})}{2} \frac{2\beta_5}{\Lambda_5 (\delta_{l_5+1})}} \sqrt{1 - \delta_{l_5}^2} \right] + e^{-\lambda_{SE} \Theta_5 - \lambda_{BS} E_1} \quad (28)$$

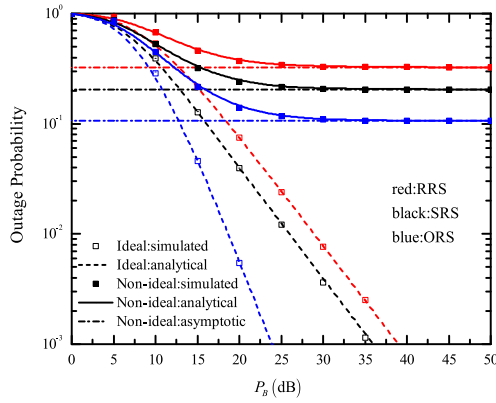


Fig. 4. OP versus the transmit power for different relay selection strategies.

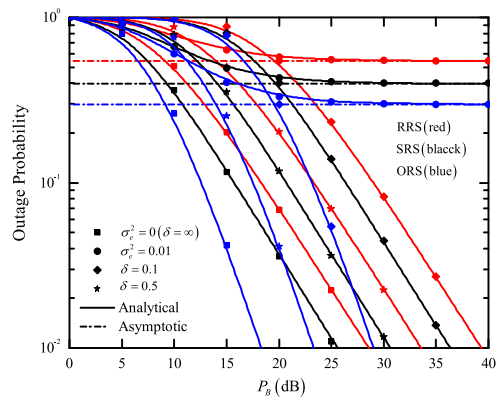


Fig. 5. OP versus the transmit power for different CEE parameters.

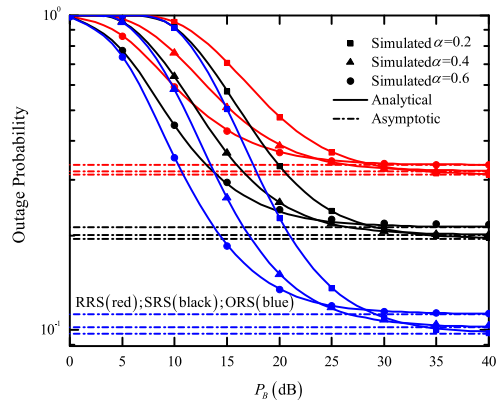


Fig. 6. OP versus the transmit power for different time allocation factors.

Fig. 5 shows the OP versus the transmit power P_B under three RS strategies for different CEE parameters. In this simulation, we set $M = 2$ [5]. The simulation results reveal that: 1) when σ_e^2 is a non-negative constant, the OPs for the three RS schemes are positively correlated with CEE parameters; 2) When $\sigma_e^2 = \Omega(1 + \delta\rho\Omega)$, the OPs decreases with the increase of δ , which means the reliability of the system increases gradually; 3) There are error floors for the OP of the three RS schemes due to fixed non-negative CEEs.

Fig. 6 shows the OP versus P_B for different time allocation efficiencies $\alpha \in \{0.2, 0.4, 0.6\}$ with $M = 2$ [40]. We have the

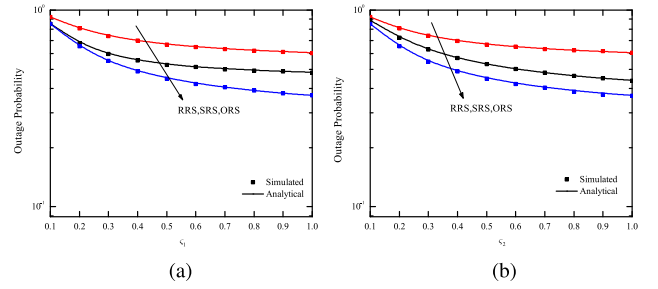


Fig. 7. Influence of energy conversion coefficient: (a) OP versus ζ_1 ; (b) OP versus ζ_2 .

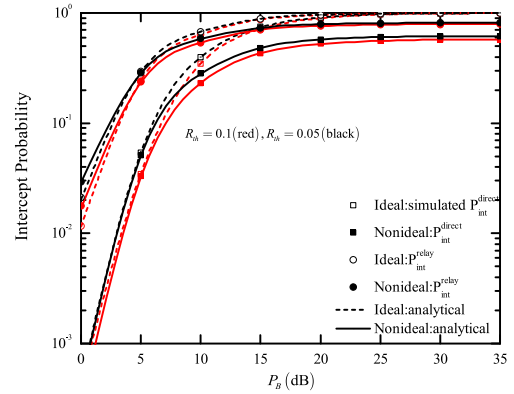


Fig. 8. IP versus the transmit power for different threshold rates and transmission schemes.

following observations that: 1) when $P_B \in [0 \text{ dB} : 24 \text{ dB}]$, the outage performance of the system becomes stronger as the α gets larger, that is, the reliability of the system increases; 2) when $P_B \in [24 \text{ dB} : 28 \text{ dB}]$, these simulation results show that OP at $\alpha = 0.6$ is higher than $\alpha = 0.4$; 3) when $P_B \in [28 \text{ dB} : 32 \text{ dB}]$, the outage performance in the case of $\alpha = 0.6$ is worse than that in $\alpha = 0.2$ and with the increase of $\alpha = 0.2; 0.4$, the OP of the system decreases; 4) when $P_B \in [32 \text{ dB} : 40 \text{ dB}]$, the OP of the system gradually weakens with $\alpha = 0.6; 0.4; 0.2$, that is, the outage performance of the system gradually improves with the changing order of time allocation efficiency.

Fig. 7(a) and Fig. 7(b) plot the OP versus energy conversion coefficients for the three RS schemes. In this simulation, the parameters are set $M = 2$, $\zeta_2 = 0.5$ in Fig. 7(a) and $\zeta_1 = 0.5$ in Fig. 7(b) [41]. From Fig. 7(a) and Fig. 7(b), we can see that the OPs under the case of the three RS schemes degrade when ζ_1 and ζ_2 grow, i.e. the reliability of the system enhances with the increase of the energy conversion coefficients of the system.

B. Security Analysis

Fig. 8 investigates the IP versus the transmit power P_B for different threshold rates and link schemes under two conditions. The parameters is set as $M = 2$ [40]. For different threshold rates, it can be seen that the IP of the system decreases as the R_{th} increases, and it can be obtained that the IP in direct link transmission condition is smaller than that in

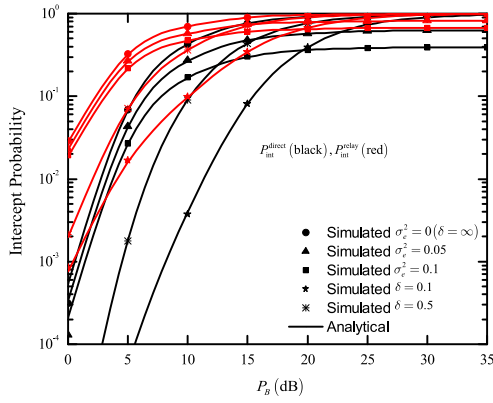


Fig. 9. IP versus the transmit power for different CEE parameters.

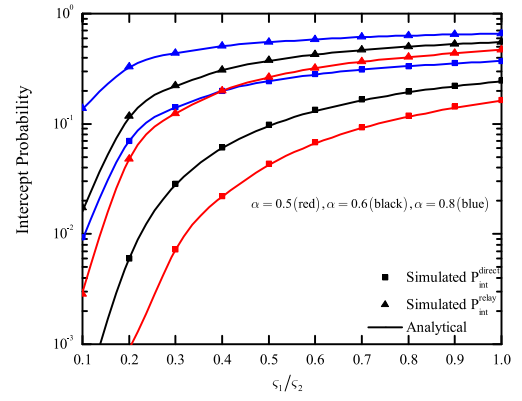


Fig. 11. IP versus the energy conversion coefficient for different time allocation factors.

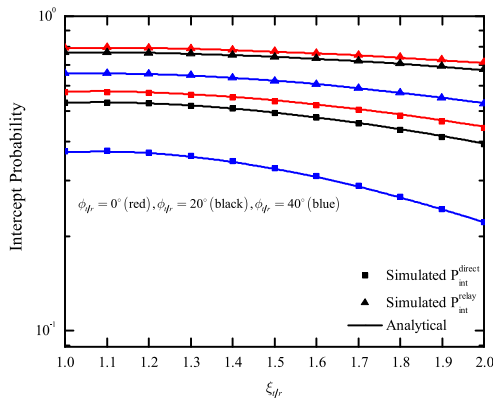


Fig. 10. IP versus the TX/RX amplitude for different phase mismatch.

relay transmission condition. This means that cooperative relay can improve the system performance by shortening the distance between source and destination. We further explore that the IP in the ideal case is smaller than the non-ideal case, that is, the presence of IQI and ICSI in the system will strengthen the security of the system in the high SNR region.

Fig. 9 illustrates the IP versus P_B for different CEE σ_e^2 . In this simulation, two CEE cases are considered: 1) σ_e^2 is a non-negative constant; 2) σ_e^2 is the function of the transmit average SNR [5]. In this simulation, we set $M = 2$. It can be seen that with the aggravation of CEE parameters, the IP of the system becomes smaller. This means that IQI parameters are beneficial to the system IP. Similarly, it can be further concluded that the IP of the system under relay transmission condition is greater than the IP under direct transmission condition.

Fig. 10 shows the IP versus TX/RX amplitude ξ_{tr} for different phase mismatch $\phi_{tr} = \{0^\circ; 20^\circ; 40^\circ\}$ [49]. The simulation results show that the IP under two transmission schemes of the system degrades with the increase of IQI parameter ξ_{tr} and phase mismatch ϕ_{tr} . This means that the IQI existing in this system is negatively correlated with IP under the conditions.

Fig. 11 plots the IP versus energy conversion coefficient ζ_{12} at S and R_m for different α with $P_B = 5$ dB [41]. From the Fig. 10, we can draw the following conclusions: 1) Under the conditions of IQI and CEEs, the IP for the two transmission schemes of the system is proportional to the time allocation

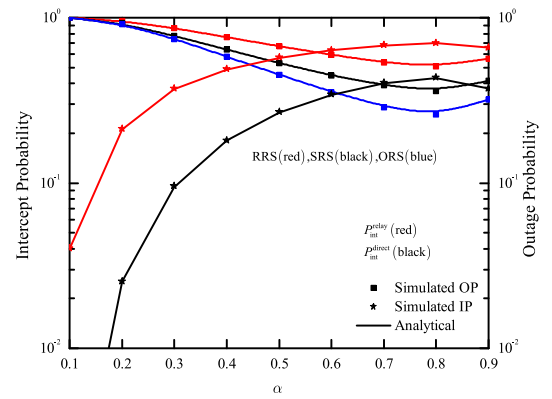


Fig. 12. OP and IP versus the time allocation factor for different transmission strategies.

factor; 2) With the increases of energy conversion coefficient, the IP under different time allocation factors gradually increases.

Fig. 12 illustrates that the OP under three RS strategies and IP under two transmission link schemes versus time allocation factor α [41]. As can be seen from the simulation, when α increases, the OP of the proposed three RS schemes decreases first and then increases, while IP under two transmission schemes increases first and then decreases in the whole range, which means that there is an optimal value in the process of α gradually increasing. In addition, the optimal solution to balance the reliability and security of the system under consideration can be obtained.

V. CONCLUSION AND FUTURE WORK

In this paper, we investigate the reliability and security of multi-relay networks in terms of OP and IP in the presence of nonlinear energy harvesters, ICSI and IQI. To improve the security performance, three RS schemes are considered. For reliability, we analyze the asymptotic behavior in the high SNR regime and discuss the diversity order. For security, we consider two representative cases. Theoretical analysis and experiment results prove that: 1) The OP of the considered system increases as the TX/RX amplitude and phase increases; 2) As

$$\varphi_2 = \frac{1}{C_1} e^{-\frac{\lambda_{SR_m} C_2}{C_1}} \left[\sqrt{\frac{\beta_1}{\gamma_1}} K_1(2\sqrt{\beta_1 \gamma_1}) - \frac{\pi \Lambda_1}{2Y_1} \sum_{l_1=0}^{Y_1} e^{-\frac{2\lambda_{BS} g_{SR_m} N_{SR_m} \varepsilon}{\Lambda_1 (\delta_{l_1} + 1)} - \frac{\Lambda_1 \lambda_{SR_m} (\delta_{l_1} + 1)}{2C_1}} \sqrt{1 - \delta_{l_1}^2} \right] \quad (\text{A.8})$$

$$Q_3 = -e^{-\lambda_{BR_m} E_2 - \lambda_{R_m D} T_3} + \frac{\lambda_{R_m D}}{C_3} e^{-\frac{\lambda_{R_m D} C_4}{C_3}} \left[\sqrt{\frac{\beta_2}{\gamma_2}} K_1(2\sqrt{\beta_2 \gamma_2}) - \frac{\pi \Lambda_2}{2Y_2} \sum_{l_2=0}^{Y_2} e^{-\frac{2\lambda_{BR_m} g_{R_m D} N_{R_m D} \varepsilon}{\Lambda_2 (\delta_{l_2} + 1)} - \frac{\lambda_{R_m D} \Lambda_2 (\delta_{l_2} + 1)}{2C_3}} \sqrt{1 - \delta_{l_2}^2} \right] \quad (\text{A.10})$$

the number of relays increases, the system's outage performance becomes better; 3) Different CEE modes have different effects on the system. When the parameter is a non-negative constant, the OP of the system increases as σ_e^2 increases. When the parameter is a variable, the OP decreases as δ increases; 4) The performance of the system is proportional to IP and energy conversion coefficient; 5) There is a trade-off between reliability and security, that is, when the performance of the interruption is relaxed, IP can be enhanced, and vice versa; 6) When the system is under the condition of nonideal and the CEEs parameter is a constant, the OP exists error floor.

The work of our paper are focusing on the secure performance of wireless-powered relaying networks affected by IQI, however, our conclusions are not specific to other hardware factors, such as, phase noise, amplifier non-linearities and quantization error, etc. To this end, the analytical method of our work can be extended to the above the hardware imperfections. In fact, the ICSI is caused not only by CCE at receiver, but also by feedback delay at the transmitter. Our analysis can be extended to investigate the secure performance of multi-antenna cooperative systems. The above exciting extensions would be done as our future work. In addition, cognitive radio technique can be introduced into our considered communication system to further analyze the security and reliability performance. Finally, the NOMA technology and relatively optimization solutions can be extend to our analysis [50].

APPENDIX A PROOF OF THEOREM 1

According to the definition of OP and (6), the following expression can be obtained as:

$$P_{out}^{RRS} = \Pr\{\min(C_{SR_m}, C_{R_m D}) < R_{th}\} \\ = 1 - \underbrace{\Pr\{C_{SR_m} > R_{th}\}}_{I_1} \underbrace{\Pr\{C_{R_m D} > R_{th}\}}_{I_2}. \quad (\text{A.1})$$

Substituting (6) into (A.1), set $\varepsilon = 2^{\frac{2R_{th}}{1-\alpha}}$ and the I_1 can be rewritten as:

$$I_1 = \Pr\left\{ \frac{|\hat{h}_{SR_m}|^2 \rho_{SR_m} p_{SR_m}}{\left[\rho_{eSR_m}^2 \rho_{SR_m} p_{SR_m} + |\hat{h}_{SR_m}|^2 \rho_{SR_m} q_{SR_m} + \sigma_{eSR_m}^2 \rho_{SR_m} q_{SR_m} + g_{SR_m} \right]} > \varepsilon \right\} \\ = Q_1 + Q_2, \quad (\text{A.2})$$

where

$$Q_1 = \Pr\left\{ \frac{g_{SR_m} N_{SR_m} \varepsilon}{C_1 |\hat{h}_{SR_m}|^2 - C_2} < |h_{BS}|^2 \leq E_1, |\hat{h}_{SR_m}|^2 \geq \frac{g_{SR_m} N_{SR_m} \varepsilon}{C_1 E_1} + \frac{C_2}{C_1} \right\}, \quad (\text{A.3})$$

and

$$Q_2 = \Pr\left\{ |\hat{h}_{SR_m}|^2 > \Theta_1, |h_{BS}|^2 > E_1 \right\} \\ = e^{-\lambda_{SR_m} \Theta_1 - \lambda_{BS} E_1}, \quad (\text{A.4})$$

in there, $T_2 = \frac{g_{SR_m} N_{SR_m} \varepsilon}{C_1 |\hat{h}_{SR_m}|^2 - C_2}$.

Substituting the PDF and CDF of Rayleigh fading into (A.3), the following formula can be obtained by further calculation as:

$$Q_1 = -\lambda_{SR_m} (e^{-\lambda_{BS} E_1} \varphi_1 - \varphi_2), \quad (\text{A.5})$$

$$\varphi_1 = \int_{T_1}^{\infty} e^{-\lambda_{SR_m} y} dy = \frac{1}{\lambda_{SR_m}} e^{-\lambda_{SR_m} T_1}, \quad (\text{A.6})$$

according to the formula (3.324.1) in [51] and the following expression (A.7) of Gaussian-Chebyshev quadrature [52], the φ_2 as shown in (A.8) at the top of the page can be obtained as

$$\int_0^{\Lambda} g(x) dx \approx \frac{\pi \Lambda}{2Y} \sum_{l=0}^Y g\left(\frac{\Lambda(\delta_l + 1)}{2}\right) \sqrt{1 - \delta_l^2}, \quad (\text{A.7})$$

Substituting (A.6) and (A.8) into (A.5), the Q_1 can be derived. Then substituting (A.4) and (A.5) into (A.2), the I_1 can be obtained. Substituting (6) into (A.1), the I_2 of following expression can be rewritten as

$$I_2 = Q_3 + Q_4 \quad (\text{A.9})$$

Similar to the calculation of I_1 , the Q_3 as shown at the top of the page and Q_4 can be obtained as follows and

$$Q_4 = e^{-\lambda_{R_m D} \Theta_2 - \lambda_{BR_m} E_2}, \quad (\text{A.11})$$

put (A.10) and (A.11) into (A.9), the I_2 can be derived.

Substituting the expressions of I_1 and I_2 , the (11) can be obtained.

APPENDIX B PROOF OF THEOREM 2

For SRS strategy, substituting (18) into (9), the following expression can be obtained as

$$\varphi_6 = \int_{T_5}^{\infty} e^{-\lambda_{SR_a}(s+1)x - \lambda_{BS}T_6} dx = \frac{1}{C_5} e^{-\frac{\lambda_{SR_a}(s+1)C_6}{C_5}} \left[\sqrt{\frac{\beta_3}{\gamma_3}} K_1 \left(2\sqrt{\beta_3\gamma_3} \right) - \frac{\pi\Lambda_3}{2Y_3} \sum_{l_3=0}^{Y_3} e^{-\frac{2\lambda_{BS}g_{SR_m}N_{SR_a}\varepsilon}{\Lambda_3(\delta_{l_3+1})} - \frac{\lambda_{SR_a}(s+1)\Lambda_3(\delta_{l_3+1})}{2C_5}} \sqrt{1 - \delta_{l_3}^2} \right] \quad (\text{B.7})$$

$$\varphi_8 = \frac{1}{C_7} e^{-\frac{\lambda_{R_aD}C_8}{C_7}} \left[\sqrt{\frac{\beta_4}{\gamma_4}} K_1 \left(2\sqrt{\beta_4\gamma_4} \right) - \frac{\pi\Lambda_4}{2Y_4} \sum_{l_4=0}^{Y_4} e^{-\frac{2\lambda_{BR_a}g_{R_mD}N_{R_aD}\varepsilon}{\Lambda_4(\delta_{l_4+1})} - \frac{\lambda_{R_aD}\Lambda_4(\delta_{l_4+1})}{2C_7}} \sqrt{1 - \delta_{l_4}^2} \right] \quad (\text{B.12})$$

$$\begin{aligned} P_{out}^{SRS} &= \Pr\{\min(C_{SR_a}, C_{R_aD}) < R_{th}\} \\ &= 1 - \underbrace{\Pr\{C_{SR_a} > R_{th}\}}_{I_3} \underbrace{\Pr\{C_{R_aD} > R_{th}\}}_{I_4}, \end{aligned} \quad (\text{B.1})$$

the CDF and PDF of $|\hat{h}_{SR_a}|^2$ can be written as

$$F_{|\hat{h}_{SR_a}|^2}(y) = [1 - e^{-\lambda_{R_aD}y}]^M, \quad (\text{B.2})$$

$$f_{|\hat{h}_{SR_a}|^2}(y) = M\lambda_{SR_a} \sum_{s=0}^{M-1} \binom{M-1}{s} (-1)^s e^{-\lambda_{SR_a}(s+1)y}. \quad (\text{B.3})$$

Similar to the calculation process of Appendix A, the I_3 and I_4 can be expressed as

$$I_3 = Q_5 + Q_6, \quad (\text{B.4})$$

where

$$\begin{aligned} Q_5 &= \Pr\{|h_{BS}|^2 (C_5 |\hat{h}_{SR_a}|^2 - C_6) > g_{SR_m} N_{SR_a} \varepsilon, |h_{BS}|^2 \leq E_1\} \\ &= \Xi(e^{-\lambda_{BS}E_1} \varphi_5 - \varphi_6), \end{aligned} \quad (\text{B.5})$$

$$\varphi_5 = \int_{T_5}^{\infty} e^{-\lambda_{SR_a}(s+1)x} dx = \frac{1}{\lambda_{SR_a}(s+1)} e^{-\lambda_{SR_a}(s+1)T_5}, \quad (\text{B.6})$$

and φ_6 as shown in (B.7) at the top of the page, and

$$\begin{aligned} Q_6 &= \Pr\{|\hat{h}_{SR_a}|^2 > \Theta_3, |h_{BS}|^2 > E_1\} \\ &= [1 - (1 - e^{-\lambda_{SR_a}\Theta_3})^M] e^{-\lambda_{BS}E_1}, \end{aligned} \quad (\text{B.8})$$

putting (B.6) and (B.7) into (B.5), the Q_5 can be obtained; substituting (B.5) and (B.8) into (B.4), the I_3 can be derived.

Then, substituting (8) into (B.1), the following formula can be expressed as

$$I_4 = Q_7 + Q_8, \quad (\text{B.9})$$

where

$$\begin{aligned} Q_7 &= \Pr\{|h_{BR_a}|^2 (C_7 |\hat{h}_{R_aD}|^2 - C_8) > g_{R_mD} N_{R_aD} \varepsilon, |h_{BR_a}|^2 \leq E_2\} \\ &= -\lambda_{R_aD} (e^{-\lambda_{BR_a}E_2} \varphi_7 - \varphi_8), \end{aligned} \quad (\text{B.10})$$

$$\varphi_7 = \int_{T_7}^{\infty} e^{-\lambda_{R_aD}y} dy = \frac{1}{\lambda_{R_aD}} e^{-\lambda_{R_aD}T_7}, \quad (\text{B.11})$$

the φ_8 as shown in (B.12) at the top of the page and Q_8 is obtained as following

$$\begin{aligned} Q_8 &= \Pr\{|\hat{h}_{R_aD}|^2 > \Theta_4, |h_{BR_a}|^2 > E_2\} \\ &= e^{-\lambda_{R_aD}\Theta_4 - \lambda_{BR_a}E_2}, \end{aligned} \quad (\text{B.13})$$

putting (B.11) and (B.12) into (B.10), the Q_7 can be obtained; substituting (B.10) and (B.13) into (B.9), the I_4 can be derived.

Substituting I_3 and I_4 into (B.1), the (19) can be obtained.

APPENDIX C

PROOF OF THEOREM 3

According to the definition of (9) and (22), the following expression for ORS strategy can be obtained as

$$\begin{aligned} P_{out}^{ORS} &= \Pr\{C_{R_m^*} < R_{th}\} \\ &= \Pr\left\{ \max_{1 \leq m \leq M} \min\{\gamma_{SR_m}, \gamma_{R_mD}\} < \varepsilon \right\} \\ &= \prod_{m=1}^M (1 - I_1 I_2), \end{aligned} \quad (\text{C.1})$$

put I_1 and I_2 of Appendix A into (C.1), the (23) can be obtained.

APPENDIX D

PROOF OF THEOREM 4

Substituting (6) into (27), the following expression can be obtained as

$$\begin{aligned} P_{int}^{direct} &\triangleq \Pr\{C_{SE} > R_{th}\} \\ &= Q_9 + Q_{10}. \end{aligned} \quad (\text{D.1})$$

Similar to the Appendix A, the Q_9 and Q_{10} can be expressed as

$$\begin{aligned} Q_9 &= \Pr\left\{ \frac{g_{SE} N_{SE} \varepsilon}{C_9 |\hat{h}_{SE}|^2 - C_{10}} < |h_{BS}|^2 \leq E_1, |h_{SE}|^2 \geq \frac{g_{SE} N_{SE} \varepsilon}{C_9 E_1} + \frac{C_{10}}{C_9} \right\} \\ &= -\lambda_{SE} (e^{-\lambda_{BS}E_1} \varphi_9 - \varphi_{10}), \end{aligned} \quad (\text{D.2})$$

$$\begin{aligned} Q_{10} &= \Pr\{|\hat{h}_{SE}|^2 > \Theta_5, |h_{BS}|^2 > E_1\} \\ &= e^{-\lambda_{SE}\Theta_5 - \lambda_{BS}E_1}, \end{aligned} \quad (\text{D.3})$$

$$\varphi_{10} = \frac{1}{C_9} e^{-\frac{\lambda_{SE} C_{10}}{C_9}} \left[2\sqrt{\frac{\beta_5}{\gamma_5}} K_1 \left(2\sqrt{\beta_5 \gamma_5} \right) - \frac{\pi \Lambda_5}{2Y_5} \sum_{l_5=0}^{Y_5} e^{-\frac{\gamma_5 \Lambda_5 (\delta_{l_5+1})}{2}} \frac{2\beta_5}{\Lambda_5 (\delta_{l_5+1})} \sqrt{1 - \delta_{l_5}^2} \right] \quad (D.5)$$

$$\varphi_{10} = \frac{1}{C_9} e^{-\frac{\lambda_{RcE} C_{10}}{C_9}} \left[\sqrt{\frac{\beta_5}{\gamma_5}} K_1 \left(2\sqrt{\beta_5 \gamma_5} \right) - \frac{\pi \Lambda_5}{2Y_5} \sum_{l_5=0}^{Y_5} e^{-\frac{2\lambda_{BRc} g_{RcE} N_{RcE} \lambda_{RcE} \Lambda_5 (\delta_{l_5+1})}{\Lambda_5 (\delta_{l_5+1}) 2C_9}} \sqrt{1 - \delta_{l_5}^2} \right] \quad (E.5)$$

where

$$\varphi_9 = \int_{T_9}^{\infty} e^{-\lambda_{SE} y} dy = \frac{1}{\lambda_{SE}} e^{-\lambda_{SE} T_9}, \quad (D.4)$$

and φ_{10} as shown at the top of the page. Putting (D.4) and (D.5) into (D.2), the Q_9 can be derived; then, substituting Q_9 and Q_{10} into (D.1), the (28) can be obtained.

APPENDIX E PROOF OF THEOREM 5

Substituting (6) into (23), the following expression can be obtained as

$$P_{\text{int}}^{\text{relay}} \triangleq \Pr\{C_{RcE} > R_{th}\} = Q_{11} + Q_{12}. \quad (E.1)$$

Similar to the Appendix A, the Q_{11} and Q_{12} can be expressed as

$$Q_{11} = \Pr \left\{ \underbrace{\frac{g_{RcE} N_{RcE}}{C_9 |\hat{h}_{RcD}|^2 - C_{10}}}_{T_{10}} < |h_{BRc}|^2 \leq E_2, \underbrace{|\hat{h}_{RcE}|^2 \geq \frac{g_{RcE} N_{RcE}}{C_9 E_2} + \frac{C_{10}}{C_9}}_{T_9} \right\} = -\lambda_{RcE} (e^{-\lambda_{BRc} E_2} \varphi_9 - \varphi_{10}), \quad (E.2)$$

$$Q_{12} = \Pr\{|\hat{h}_{RcE}|^2 > \Theta_5, |h_{BRc}|^2 > E_2\} = e^{-\lambda_{RcE} \Theta_5 - \lambda_{BRc} E_2}, \quad (E.3)$$

where

$$\varphi_9 = \int_{T_9}^{\infty} e^{-\lambda_{RcE} y} dy = \frac{1}{\lambda_{RcE}} e^{-\lambda_{RcE} T_9}, \quad (E.4)$$

and φ_{10} as shown in (E.5) at the top of the page.

Putting (E.4) and (E.5) into (E.2), the Q_{11} can be derived; then, substituting Q_{11} and Q_{12} into (E.1), the (29) can be obtained.

REFERENCES

- [1] S. Jacob *et al.*, "A novel spectrum sharing scheme using dynamic long short-term memory with CP-OFDMA in 5G networks," *IEEE Trans. Cognitive Commun. Netw.*, vol. 6, no. 3, pp. 926–934, Sep. 2020.
- [2] Y. Liu, H. Chen, and L. Wang, "Physical layer security for next generation wireless networks: Theories, technologies, and challenges," *IEEE Commun. Surv. Tut.*, vol. 19, no. 1, pp. 347–376, Jan.–Mar. 2017.
- [3] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surv. Tut.*, vol. 16, no. 3, pp. 1550–1573, Jul.–Sep. 2014.
- [4] X. Li *et al.*, "Hardware impaired ambient backscatter NOMA systems: Reliability and security," 2020. [Online]. Available: <https://arxiv.org/abs/2008.05798>
- [5] X. Li, M. Huang, J. Li, Q. Yu, K. Rabie, and C. C. Cavalcante, "Secure analysis of multi-antenna cooperative networks with residual transceiver HIs and CEEs," *IET Commun.*, vol. 13, no. 17, pp. 2649–2659, Oct. 2019.
- [6] M. L. Ammari and P. Fortier, "Physical layer security of multiple-input multiple-output systems with transmit beamforming in rayleigh fading," *IET Commun.*, vol. 9, no. 8, pp. 1096–1103, 2015.
- [7] H. Lei, C. Gao, Y. Guo, and G. Pan, "On physical layer security over generalized gamma fading channels," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1257–1260, Jul. 2015.
- [8] J. Sun, X. Li, M. Huang, Y. Ding, J. Jin, and G. Pan, "Performance analysis of physical layer security over $\kappa - \mu$ shadowed fading channels," *IET Commun.*, vol. 12, no. 8, pp. 970–975, 2018.
- [9] Y. Zou, X. Wang, and W. Shen, "Optimal relay selection for physical-layer security in cooperative wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2099–2111, Oct. 2013.
- [10] S. Parkvall, A. Furusker, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 84–91, Feb. 2011.
- [11] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, C. Koutsimanis, and X. Shen, "Relaying operation in 3GPP LTE: Challenges and solutions," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 156–162, Feb. 2012.
- [12] L. Xu, J. Wang, H. Zhang, and T. A. Gulliver, "Performance analysis of IAF relaying mobile d2d cooperative networks," *J. Franklin Inst.*, vol. 354, no. 2, pp. 902–916, Oct. 2017.
- [13] L. Xu, J. Wang, H. Wang, T. A. Gulliver, and K. N. Le, "BP neural network-based ABEP performance prediction for mobile internet of things communication systems," *Neural Comput. Appl.*, vol. 32, pp. 16025–16041, 2020.
- [14] A. Behnad, N. C. Beaulieu, and B. Maham, "Multi-hop amplify-and-forward relaying on nakagami-0.5 fading channels," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 173–176, Jun. 2012.
- [15] Y. Liu and A. P. Petropulu, "Destination assisted cooperative jamming for wireless physical layer security," in *Proc. IEEE Int. Workshop Inf. Forensics Security*, Dec. 2012, pp. 282–287.
- [16] P. N. Son, V. P. Tuan, S. Park, and H. Y. Kong, "Closed-form analysis of a decode-and-forward scheme under physical layer security over general fading channels," in *Proc. 5th NAFOSTED Conf. Inform. Comput. Sci.*, Nov. 2018, pp. 1–5.
- [17] H. U. Sokun and H. Yanikomeroglu, "On the spectral efficiency of selective decode-and-forward relaying," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4500–4506, May 2017.
- [18] Y. Zhao, R. Adve, and T. J. Lim, "Symbol error rate of selection amplify-and-forward relay systems," *IEEE Commun. Lett.*, vol. 10, no. 11, pp. 757–759, Nov. 2006.
- [19] S. M. I. Krikidis, J. Thompson and N. Goertz, "Amplify-and-forward with partial relay selection," *IEEE Commun. Lett.*, vol. 12, no. 4, pp. 235–237, Apr. 2008.
- [20] H. Lei *et al.*, "On secrecy outage of relay selection in underlay cognitive radio networks over nakagami- m fading channels," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 614–627, Dec. 2017.
- [21] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.
- [22] S. Bernard and J. Fredric, *Digital Communications: Fundamentals and Applications*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice Hall, 1988.
- [23] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Commun. Sur.Tut.*, vol. 13, no. 3, pp. 443–461, Jul–Sep 2011.

- [24] M. Ku, W. Li, Y. Chen, and K. J. Ray Liu, "Advances in energy harvesting communications: Past, present, and future challenges," *IEEE Commun. Sur. Tut.*, vol. 18, no. 2, pp. 1384–1412, Apr–Jun 2016.
- [25] Z. Ding *et al.*, "Application of smart antenna technologies in simultaneous wireless information and power transfer," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86–93, Apr. 2015.
- [26] I. Krikidis, S. Timotheou, S. Nikolaou, G. Zheng, D. W. K. Ng, and R. Schober, "Simultaneous wireless information and power transfer in modern communication systems," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 104–110, Nov. 2014.
- [27] H. Xing, L. Liu, and R. Zhang, "Secrecy wireless information and power transfer in fading wiretap channel," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 180–190, Jan. 2016.
- [28] J. Zhang, G. Pan, and H. Wang, "On physical-layer security in underlay cognitive radio networks with full-duplex wireless-powered secondary system," *IEEE Access*, vol. 4, pp. 3887–3893, 2016.
- [29] F. Jameel, W. U. Khan, N. Kumar, and R. Jntti, "Efficient power-splitting and resource allocation for cellular V2X communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3547–3556, Jun. 2021.
- [30] F. Jameel, S. Wyne, and Z. Ding, "Secure communications in three-step two-way energy harvesting DF relaying," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 308–311, Feb. 2018.
- [31] E. Boshkovska, D. W. K. Ng, N. Zlatanov, A. Koelpin, and R. Schober, "Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 1984–1999, May 2017.
- [32] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical nonlinear energy harvesting model and resource allocation for SWIPT systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082–2085, Dec. 2015.
- [33] Y. Dong, M. J. Hossain, and J. Cheng, "Performance of wireless powered amplify and forward relaying over nakagami- m fading channels with nonlinear energy harvester," *IEEE Commun. Lett.*, vol. 20, no. 4, pp. 672–675, Apr. 2016.
- [34] X. Li, J. Li, Y. Liu, Z. Ding, and A. Nallanathan, "Residual transceiver hardware impairments on cooperative NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 680–695, Jan. 2020.
- [35] S. Bernard, "Digital communications: fundamentals and applications," *Englewood Cliffs, NJ, USA: Prentice Hall*, 2001.
- [36] S. Mirabbasi and K. Martin, "Classical and modern receiver architectures," *IEEE Commun. Mag.*, vol. 38, no. 11, pp. 132–139, Nov. 2000.
- [37] Ö. Özdemir, R. Hamila, and N. Al-Dhahir, "Exact average OFDM sub-carrier SINR analysis under joint transmit-receive I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 4125–4130, Oct. 2014.
- [38] X. Li *et al.*, "Physical layer security of cooperative noma for IoT networks under I/Q imbalance," *IEEE Access*, vol. 8, pp. 51 189–51 199, 2020.
- [39] X. Li, M. Zhao, Y. Liu, L. Li, Z. Ding, and A. Nallanathan, "Secrecy analysis of ambient backscatter NOMA systems under I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12286–12290, Oct. 2020.
- [40] X. Li *et al.*, "Security and reliability performance analysis of cooperative multi-relay systems with nonlinear energy harvesters and hardware impairments," *IEEE Access*, vol. 7, pp. 102 644–102 661, 2019.
- [41] J. Zhang, G. Pan, and Y. Xie, "Secrecy analysis of wireless-powered multi-antenna relaying system with nonlinear energy harvesters and imperfect CSI," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 460–470, Jun. 2018.
- [42] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical nonlinear energy harvesting model and resource allocation for SWIPT systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082–2085, Dec. 2015.
- [43] X. Li, J. Li, P. T. Mathiopoulos, D. Zhang, L. Li, and J. Jin, "Joint impact of hardware impairments and imperfect CSI on cooperative SWIPT NOMA multi-relaying systems," in *Proc. IEEE/CIC Int. Conf. Commun. China*, Aug. 2018, pp. 95–99.
- [44] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [45] S. Tim, *RF Imperfections in High-Rate Wireless Systems: Impact and Digital Compensation*. Berlin, Germany: Springer, 2008.
- [46] J. Qi, S. Aissa, and M. Alouini, "Analysis and compensation of I/Q imbalance in amplify-and-forward cooperative systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2012, pp. 215–220.
- [47] X. Li *et al.*, "Security analysis of multi-antenna NOMA networks under I/Q imbalance," *Electronics*, vol. 8, no. 11, pp. 1–17, Nov. 2019.
- [48] J. Li, M. Matthaiou, and T. Svensson, "I/Q imbalance in AF dual-hop relaying: Performance analysis in nakagami- m fading," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 836–847, Mar. 2014.
- [49] R. Hamila, Ö. Özdemir, and N. Al-Dhahir, "Beamforming OFDM performance under joint phase noise and I/Q imbalance," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2978–2989, May 2016.
- [50] W. U. Khan, F. Jameel, T. Ristaniemi, S. Khan, G. A. S. Sidhu, and J. Liu, "Joint spectral and energy efficiency optimization for downlink NOMA networks," *IEEE Trans. Cognitive Commun. Netw.*, vol. 6, no. 2, pp. 645–656, Jun. 2020.
- [51] S. G. Izrail and M. R. Iosif, *Table of Integrals, Series, and Products*. New York, NY, USA: Academic, 2014.
- [52] B. H. Francis, *Introduction to Numerical Analysis*. Chelmsford, MA, USA: Courier Corporation, 1987.

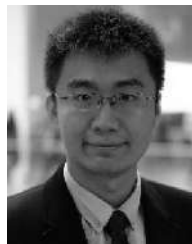


Xingwang Li (Senior Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2010, and the Ph.D. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2015. From 2010 to 2012, he was working with Comba Telecom, Ltd., Guangzhou, China, as an Engineer. He spent one year from 2017 to 2018 as a visiting scholar with Queen's University Belfast, Belfast, U.K. He is also a visiting scholar with State Key Laboratory of Networking and Switching Technology, Beijing

University of Posts and Telecommunications from 2016 to 2018. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. His research interests include MIMO communication, cooperative communication, hardware constrained communication, non-orthogonal multiple access, physical layer security, unmanned aerial vehicles, and the Internet of Things. He has served as many TPC members, such as the IEEE/CIC, GLOBECOM, WCNC, and so on. He serves as the Co-Chair for the IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track. He also serves as an Editor on the Editorial Board for the IEEE *ACCESS*, *Computer Communications*, and *KSII Transactions on Internet and Information Systems*. He is also the Lead Guest Editor for the Special Issue on Recent Advances in Physical Layer Technologies for the 5G-Enabled Internet of Things of Wireless Communications and Mobile Computing and Special Issue on Recent Advances in Multiple Access for 5G-enabled IoT of Security and Communication Networks.



Mengyan Huang (Student Member, IEEE) received the B.Sc. degree in communication engineering (wireless mobile communication direction) from the College of Electronic Information Engineering, Sias International College of Zhengzhou University, Zhengzhou, China, in 2017. She is currently working toward the M.Sc. degree in communication and information systems with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, China. Her current research interests include massive MIMO and physical layer secure communication.



Yuanwei Liu (Senior Member, IEEE) received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011 and 2014, respectively, and the Ph.D. degree in electrical engineering from the Queen Mary University of London, London, U.K., in 2016. He was with the Department of Informatics, King's College London, from 2016 to 2017, where he was a Postdoctoral Research Fellow. Since 2017, he has been a Lecturer (Assistant Professor) with the School of Electronic Engineering and Computer Science, Queen Mary

University of London. His current research interests include 5G wireless networks, the Internet of Things, machine learning, stochastic geometry, and matching theory. He has served as a TPC Member for many IEEE conferences, such as GLOBECOM and ICC. He received the Exemplary Reviewer Certificate for the IEEE WIRELESS COMMUNICATION LETTERS in 2015, the IEEE TRANSACTIONS ON COMMUNICATIONS in 2016, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2017. He currently serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE COMMUNICATIONS LETTERS, and IEEE ACCESS. He is also a Guest Editor for the IEEE JSTSP special issue on Signal Processing Advances for Non-Orthogonal Multiple Access in Next-Generation Wireless Networks.



Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, School of Communication and Management Studies School of Engineering and Technology, Kochi, India. He is a Distinguished Speaker of ACM. He has authored or coauthored more than 50 research papers in peer reviewed and highly indexed international journals and conferences. He has served more than 20

conferences, such as IEEE ICC, ICCCN 2020, IEEE COINS 2020, SigTelCom, ICACCI, ICDMAI in leadership capacities including program Co-Chair, track Chair, session Chair, and Technical Program Committee Member. His research interests include Internet of Things, fog computing and networking, underwater acoustic sensor networks, scientometrics, educational psychology, ad-hoc networks, wireless communication, opportunistic routing, and wireless sensor networks. He is currently a Guest Editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE SENSORS JOURNAL, and IEEE INTERNET OF THINGS JOURNAL. He is an Associate Editor for *IET Quantum Communications* and also an Editorial Board Member of IEEE Future Directions.



Anand Paul (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently an Associate Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea. He is a delegate representing South Korea for M2M focus group and for MPEG. His research interests include algorithm and architecture reconfigurable embedded computing. He has guest edited various international journals and he is also part of Editorial Team for *Journal of Platform*

Technology, *ACM Applied Computing Review*, and *Cyber-Physical Systems*. He serves as a Reviewer for various IEEE /IET/Springer and Elsevier journals. He is the track chair for Smart human computer interaction in ACMSAC 2015, 2014. He was the recipient of the Outstanding International Student Scholarship Award in 2004–2010, the Best Paper Award in National Computer Symposium and in 2009, and International Conference on Soft computing and Network Security, India, in 2015.



Zhiguo Ding (Fellow, IEEE) received the B.Eng. degree in electrical engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2000, and the Ph.D. degree in electrical engineering from Imperial College London, London, U.K., in 2005. From July 2005 to April 2018, he was working with Queen's University Belfast, Imperial College, Newcastle University and Lancaster University. Since April 2018, he has been with the University of Manchester, Manchester, U.K., as a Professor in communications. From October 2012 to September 2020, he has

also been an Academic Visitor with Princeton University. His research interests include 5G networks, game theory, cooperative and energy harvesting networks, and statistical signal processing. He is serving as an Area Editor for the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and *Journal of Wireless Communications and Mobile Computing*, and was an Editor for the IEEE WIRELESS COMMUNICATION LETTERS, IEEE COMMUNICATION LETTERS from 2013 to 2016. He received the Best Paper Award in IET ICWMC-2009 and IEEE WCSP-2014, the EU Marie Curie Fellowship 2012–2014, the Top IEEE TVT Editor 2017, IEEE Heinrich Hertz Award 2018, IEEE Jack Neubauer Memorial Award 2018, IEEE Best Signal Processing Letter Award 2018, and Web of Science Highly Cited Researcher 2019.



High-performance flow classification using hybrid clusters in software defined mobile edge computing

Mahdi Abbasi^a  , Azad Shokrollahi^a, Mohammad R. Khosravi^{b c}, **Varun G. Menon^d**

^a Department of Computer Engineering, Engineering Faculty, Bu-Ali Sina University, Hamedan, Iran

^b Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

^c Telecommunications Group, Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran

^d Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683582, Kerala, India

Received 24 February 2020, Revised 17 June 2020, Accepted 1 July 2020, Available online 6 July 2020, Version of Record 13 July 2020.



Show less 

 Share  Cite

<https://doi.org/10.1016/j.comcom.2020.07.002> 

[Get rights and content](#) 

Abstract

Mobile Edge Computing (MEC) provides different storage and computing capabilities within the access range of mobile devices. This moderates the burden of offloading compute/storage-intensive processes of the mobile devices to the centralized cloud data centers. As a result, the network latency is reduced and the quality of service provided for the mobile end users is improved. Different applications benefit from the large-scale deployments of MEC servers. However, the considerable complexity of managing large scale deployments of the sheer number of applications for the millions of mobile devices is a challenge. Recently, Software Defined Networking (SDN) is leveraged to resolve the problem by providing unified and programmable interfaces for managing network devices. Most of the current SDN packet processing services are tightly dependent on the packet classification service. This primary service classifies any incoming packet based on matching a set of specific fields of its header against a flow table. Acceleration of this basic process considerably increases the performance of the SDN-based MEC. In this paper, the hierarchical tree algorithm, which is a packet classification method, is parallelized using popular platforms on a cluster of Graphics Processing Units (GPUs), a cluster of Central Processing Units (CPUs), and a hybrid cluster. The best scenario for the parallel implementation of this algorithm on the CPU cluster is that which combines OpenMP and MPI.

In this case, the throughput of the classifier is 4.2 million packets per second (MPPS). On the GPU cluster, two different scenarios have been used. In the first scenario, the global memory is used to store the rules and the Hierarchical-trie of the classifier while in the second scenario we break the filter set in a way that the resulting Hierarchical-trie of each subset could be stored in the shared memory of GPU. According to the results, although the first GPU cluster scenario achieves a throughput of 29.19 MPPS and a speedup 58 times as great as the serial mode, the second scenario is 12

times faster due to using the shared memory. The best performance, however, belongs to the hybrid cluster mode. The hybrid cluster achieves a throughput of 30.59 which is 1.4 MPPS more than the GPU cluster.

Introduction

The advent of several mobile applications, such as intelligent transport systems[1], virtual reality[2], Human activity recognition, and control[3], [4], [5], [6], and smart environments[1], [7], [8] has implied the availability of a large pool of computing and storage resources. Hence, the considerable growth in data volume that comes from the massive number of devices enabled by 5G has made mobile edge computing more important than ever before[9], [10]. Beyond its abilities to reduce network traffic and improve user experience, edge computing also plays a critical role in enabling use cases for ultra-reliable low-latency communication in industrial manufacturing and a variety of other sectors[11]. Especially, facilitating cloud-like resources at the edge of the network is a challenging issue for the telecommunication sector[12]. By the introduction of the Mobile Edge Computing (MEC) in 2014, a sustainable business model is provided for mobile operators, service providers, and mobile subscribers[13], [14]. MEC aims to provide cloud capabilities within the Radio Access Network (RAN) in the area of mobile subscribers[15]. That is, it provides accelerated services, contents, and applications by increasing availability at the edge[16]. Recently, flexible and scalable solutions of SDN for a large set of challenges that are encountered within the traditional networking approach, have introduced it as a suitable collaborator for MEC[17]. When the intrinsic properties of SDN are considered, three practical models for fruitful collaboration of SDN and MEC in real-world scenarios would be exploited, including multi-tier edge computing architecture, service-centric access to the edge, and network function virtualization (NFV)[17], [18]. SDN has proficiencies of arranging the network, its services, and devices by hiding the complexities of the varied mobile environment from end-users. Thus, SDN can moderate the barriers and limits that multi-tier MEC infrastructure will meet. The SDN control mechanism can reduce the complexity of MEC by utilizing accessible resources more efficiently. SDN dynamically routes the traffic between MEC servers and cloud servers to deliver the highest quality of service to end-users. In the second model of collaboration, the NFV platform of SDN can be dedicated to MEC or shared with other network functions or applications. In this model, MEC can use NFV management and orchestration entities and interfaces. Finally, as the third form of collaboration, SDN provides high speeds in content delivery between the MEC and central cloud systems.

These collaboration models result in several benefits including high resolution & effective control, flexibility and low barrier on innovation, service-centric implementation, virtual machine mobility, adaptability, interoperability, low-cost solutions, and multiplicity of scope[17].

SDN uses the limited network resources optimally and enables flexible network management by separating the control operations from the data management[19]. For this purpose, the forwarding switch nodes cannot take decisions on their own, instead, a software-based controller that has a general view of the underlying network makes the forwarding and routing decisions. All operations of the SDN controller, especially its flexible communication with switches are carried out according to the OpenFlow protocol. The chief functionalities of the SDN controller include managing the flow tables on the forwarding nodes, collecting statistics, and populating them by editing the packet classification rules. To classify a packet appearing at an ingress port of an SDN switch, the switch performs a lookup on its flow table, according to a classification algorithm to find any matching rule. If found, the corresponding action on the packet. Otherwise, the switch requests the controller to figure out the most appropriate action. The decided action is applied to the packet, and the necessary classification rule is installed on the corresponding switch.

The performance of the classification engine of an SDN switch has a great impact on the overall performance of the system[20], [21]. There are different methods for parallel implementation of packet classification algorithms[1], [22], [23], [24], [25], [26]. Some of these methods, e.g. parallelization of algorithms on GPUs and multi-core CPUs, have been recently implemented. Due to the limitation of hardware resources, however, the throughput of these systems can hardly reach the cumulative throughput rates of current network systems. A common solution to overcome this limitation is CPU clusters and GPU clusters. The present study is the first attempt to parallelize the Hierarchical-trie algorithm on a CPU cluster, a GPU cluster, and a hybrid cluster. The innovations of this study are as follows:

- For the first time, a cluster system has been used for packet classification.

- This study compares for the first time the performance of programming based on Message Passing Interface (MPI) with that of OpenMP-based programming in cluster computations.
- In all the scenarios, the effect of different types of memory in the hierarchy of GPU memory on the performance of concurrent processes of packet classification in a GPU cluster is investigated.
- The parameters of memory usage, speedup, and throughput are measured based on the results of our implementation of the scenarios which are combinations of MPI, CUDA, and OpenMP.

The paper is organized as follows. Section2 discusses the Hierarchical-trie packet classification algorithm. Next, cluster systems and their programming will be described. In Section3, the literature on different architectures of the cluster and parallel implementations of packet classification algorithms will be reviewed. Section4 provides a description of the proposed scenarios for the parallelization of Hierarchical-trie algorithm on CPU clusters, GPU clusters, and hybrid clusters. In the next section, the implementation results will be analyzed and evaluated. The final section compares the results of our work with other findings in this field and proposes suggestions for further research.

Section snippets

Tools and algorithms

This section describes the structure of the Hierarchical-trie algorithm as well as how this algorithm classifies internet packets. Next, cluster computing and its related parallelization tools are discussed....

Review of literature

Zhou et al. conducted one of the preliminary research studies of the parallelization of packet classification algorithms on multi-core systems[43]. They parallelized linear search algorithm and area-based tree search algorithm using Pthread library. The maximum throughput of their parallel packet classifier was 11.5 Gbps. Another dominant study was conducted by Qu et al. in 2015. They parallelized the Bit-vector packet classification algorithm on multi-core processors using the OpenMP. The...

Proposed method

In this section, we shall first describe the cluster used for the implementation. Next, we shall explain the parallelization of Hierarchical-trie algorithm on a CPU cluster, a GPU cluster, and a hybrid cluster. As Table2 predicts, the hybrid cluster can reveal the highest performance level. Also, it is expected that the complexity of a hybrid cluster algorithm is more than that of any parallel algorithm.

...

Implementation and evaluation

In this section, we will first describe a software suite for generating the filter sets of experimental headers. Next, we will discuss our criteria and evaluate the results from the scenarios described in Section4....

Conclusion

MEC is a new accelerated trend in ubiquitous computing where the computational resources are being brought nearer to the mobile users. SDN, can serve as an enabler to reduce the complexity barriers involved and let the real potential of MEC be achieved. That is, the complexities resulted from deploying the cloud-like resources and related services at the edge of the mobile network can be solved by a control mechanism that can orchestrate the distributed environment. All data flow management,...

CRedit authorship contribution statement

Mahdi Abbasi: Supervision, Conceptualization, Methodology, Formal analysis, Visualization, Writing - review & editing. **Azad Shokrollahi:** Data curation, Software, Writing - original draft. **Mohammad R. Khosravi:** Conceptualization, Validation, Writing - review & editing. **Varun G. Menon:** Methodology, Validation, Writing - review & editing....

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper....

[Special issue articles](#) [Recommended articles](#)

References (52)

WanS. *et al.*

[Cognitive computing and wireless communications on the edge for healthcare service robots](#)

Comput. Commun. (2020)

WanS. *et al.*

[Faster R-CNN for multi-class fruit detection using a robotic vision system](#)

Comput. Netw. (2020)

JiangC. *et al.*

[Energy aware edge computing: A survey](#)

Comput. Commun. (2020)

WanS. *et al.*

[Multi-dimensional data indexing and range query processing via Voronoi diagram for internet of things](#)

Future Gener. Comput. Syst. (2019)

SaraswatS. *et al.*

[Challenges and solutions in software defined networking: A survey](#)

J. Netw. Comput. Appl. (2019)

LinF. *et al.*

[High-performance IPv6 address lookup in GPU-accelerated software routers](#)

J. Netw. Comput. Appl. (2016)

GongT. *et al.*

[GPU-based parallel optimization of immune convolutional neural network and embedded system](#)

Eng. Appl. Artif. Intell. (2017)

GhidoucheK. *et al.*

[Efficient high degree polynomial root finding using GPU](#)

J. Comput. Sci. (2017)

FernándezJ.L. *et al.*

[GPU parallel implementation for asset-liability management in insurance companies](#)

J. Comput. Sci. (2018)

AbbasiM. *et al.*

An efficient parallel genetic algorithm solution for vehicle routing problem in cloud implementation of the intelligent transportation systems

J. Cloud Comput. (2020)



View more references

Cited by (22)

[Multi-view clustering via matrix factorization assisted k-means](#)

2023, Neurocomputing

[Show abstract](#)

[Reliability and robust resource allocation for Cache-enabled HetNets: QoS-aware mobile edge computing](#)

2022, Reliability Engineering and System Safety

Citation Excerpt :

...According to the previous studies, some frameworks have been developed for dynamic power optimization in cache-enabled systems and power-sharing, respectively [9–11]. The aforementioned works in [12] are mainly concentrated on the performance analysis, rate maximization, power minimization problem for NOMA-based mobile edge computing and cooperative multi-server cloud systems. At present, worldwide is advocating wireless technologies as backhauling to obtain green communication, aiming to increase energy efficiency, and providing higher bit rate services....

[Show abstract](#)

[Special Issue on Optimization of Cross-layer Collaborative Resource Allocation for Mobile Edge Computing, Caching and Communication](#)

2022, Computer Communications

[Intelligent workload allocation in IoT–Fog–cloud architecture towards mobile edge computing](#)

2021, Computer Communications

Citation Excerpt :

...Also, AT&T networks use 200 petabytes of bandwidth each year. Sending all the data to the cloud requires a large amount of bandwidth used for sophisticated methods of high-performance flow processing [13,17,18]. Many IoT devices have limited resources and are unable to fulfill their own computation needs [8,15]....

[Show abstract](#)

[ABM-SpConv-SIMD: Accelerating Convolutional Neural Network Inference for Industrial IoT Applications on Edge Devices](#)

2023, IEEE Transactions on Network Science and Engineering

[A Highly Compatible Verification Framework with Minimal Upgrades to Secure an Existing Edge Network](#)



2023, ACM Transactions on Internet Technology



View all citing articles on Scopus

[View full text](#)



All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.



Enhancing the Performance of Flow Classification in SDN-Based Intelligent Vehicular Networks

Mahdi Abbasi¹, Hajar Rezaei, Varun G. Menon², *Senior Member, IEEE*,
Lianyong Qi³, *Member, IEEE*, and Mohammad R. Khosravi⁴

Abstract—Intelligent vehicular networks converged with software-defined networking provides several flow-based surveillance services to mobile applications on vehicular nodes. But, as the scale of such networks grows exponentially, a substantial delay in processing tremendous flows emerges. The delay can be reduced by accelerating the packet classification methods, which are nowadays exploited in software-defined vehicular networks. Fast packet classification lets firewalls to inspect each incoming packet at wire speed. One of the well-known packet classification methods is the KD-tree algorithm. This paper presents an enhanced version of this algorithm that uses the geometric space to display different fields and increases search speed by recursive decomposition of the search space. Also, the enhanced KD-tree is integrated with a leaf-pushing technique, which enhances the performance of KD-tree search during classification. The proposed algorithm is implemented using a bloom filter data structure and a hash table. Experimental results show that the proposed leaf-pushed KD-tree algorithm improves packet classification speed up to 24 times in comparison with the conventional KD-tree. Moreover, the proposed algorithm can significantly reduce the classification time in comparison with state-of-the-art tree-based algorithms.

Index Terms—Intelligent vehicular network, flow classification, KD-tree algorithm, leaf-pushing, performance, software-defined-networking (SDN).

I. INTRODUCTION

INTELLIGENT Vehicular Network (IVN) is one of the world-evolving technologies that help enhance road safety and efficient traffic control in smart cities [1]. This technology uses various communication technologies to provide organized routes to high mobility vehicular nodes [2], [3]. Although recently exploited high-speed communication technologies can provide dependable and universal mobile coverage [4], several

prominent features of novel deployments of IVN lead new challenges, such as unbalanced traffic flow in a multi-path topology and inefficient network utilization [5]–[7]. Thus, flexible and programmable architectures like software-defined-networking (SDN) have been recently proposed as a key solution for IVNs. The network programmability feature of the SDN, when added to IVN lets external applications to simply reconfigure the equipment and wireless devices [8]. That is, the SDN provides considerable flexibility in evolving vehicular network infrastructure [9], [10]. For this purpose, flow classification rules are configured and assigned to switches dynamically according to the network conditions and the requirements for applications on IVN [11]. Flow classification enables an SDN controller to provide several on-demand IVN surveillance services. Each SDN controller manages a dynamic set of packet classification rules, each of which corresponds to a data stream to/from a specific vehicular node [11], [12]. An essential prerequisite for classifying data into specific flows is the packet classification [13]–[17]. Packet classification refers to the process of classifying network packets into flows in routers and switches. Various methods have been so far developed for this purpose which are different in terms of classification time and memory usage. The methods are either software-based or hardware-based. Major hardware-based methods make use of Field-Programmable Gate Array (FPGA) and Ternary Content-Addressable Memory (TCAM) [18]. In general, although hardware-based classifiers achieve high speeds and throughput rates up to 100 MPPS (million packets per second), they cannot be easily developed and customized due to the limited resources on the chip [19], [20]. Moreover, these systems carry high costs and have a low efficiency-to-cost ratio. This is why software-based methods have become the focus of attention in recent years [19]–[24].

In spite of their extensibility, software-based classifiers do not function efficiently in networks with high bandwidth due to the low speed of the serial processing of instructions in CPUs. The challenge of accelerating the software-based classifiers of IP packets, therefore, has resulted in considerable research with the aim of developing methods to increase the speed of classification algorithms. In this study, we seek to use the leaf pushing technique to enhance the performance of KD-trees. The KD-tree is a decision-tree packet classification algorithm. Decision tree-based algorithms are considered as an important class of software-based classification methods. In this type of classification, the rule sets are stored in the search tree based

Manuscript received February 28, 2020; revised July 8, 2020; accepted July 30, 2020. Date of publication August 13, 2020; date of current version July 12, 2021. This work was supported by Bu-Ali Sina University. The Associate Editor for this article was S. Mumtaz. (*Corresponding author: Mahdi Abbasi.*)

Mahdi Abbasi and Hajar Rezaei are with the Department of Computer Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan 6516738695, Iran (e-mail: abbasi@basu.ac.ir; rezaei@eng.basu.ir).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683582, India (e-mail: varunmenon@scmsgroup.org).

Lianyong Qi is with the School of Information Science and Engineering, Qufu Normal University, Jining 273165, China (e-mail: lianyongqi@gmail.com).

Mohammad R. Khosravi is with the Department of Computer Engineering, Persian Gulf University, Bushehr 75169-13817, Iran, and also with the Telecommunications Group, Shiraz University of Technology, Shiraz 71555-313, Iran (e-mail: mohammadkhosravi@acm.org).

Digital Object Identifier 10.1109/TITS.2020.3014044

on binary patterns in the rule fields. Hence, to find the rule that best matches the incoming packet, the tree is traversed based on the binary content of the fields in question [25]. Various tree-based algorithms such as AQT [26], HiCuts [27], and Hyper-Cuts [28] have been so far developed. These algorithms first, seek to obtain efficient search methods by using the geometric representation of rules, and then construct the corresponding decision tree.

As the main contribution, we propose a classification method that makes use of leaf-pushing to allocate search space in a KD-tree. In this method, the nodes in each path that contain rules are reduced to one leaf node. The rules to be compared with each packet are confined to the rules stored in the leaf node and the process of searching the tree is completely separated from the process of rule comparison. As a result of optimizing the KD-tree and using leaf pushing technique, both memory usage and access to off-chip memory are reduced.

The paper is organized as follows. Section II reviews the related literature. Next, the proposed method is described in Section III and evaluated in Section IV. The final section concludes the discussion and shows the direction of further research.

II. RELATED WORK

In this section, the Area-based Quad Tree (AQT) algorithm and the other relevant methods are briefly explained. Next, the key idea behind leaf-pushing is fully explained.

A. Area-Based Quad Tree

In this algorithm, each packet is represented as a point in the geometric space. Space decomposition algorithms provide a search technique that uses a tree or tree-like structure to find a rule that covers the packet. An area-based quad tree (AQT) has a search area that consists of the source prefix address on the X axis and the destination prefix address on the Y axis. Each rule is represented as a square formed by the source and destination prefix addresses [26].

B. Other Algorithms

Linear search compares the rules sequentially with the incoming packet and has a low performance in terms of time. Characteristic of space decomposition algorithms is their geometric approach. In fact, the space of the classification problem is represented as a d-dimensional geometric space in which separators are shown as rectangles. While the rules are stored only once in AQT, other space decomposition algorithms allow their repetition to increase the efficiency of packet classification. Hierarchical Intelligent Cutting algorithm (HiCuts), for example, produces a decision tree by recursive decomposition of the search space. On each node of the tree, one decision is applied to decompose the current search space into several subsets so that each subset would specify a child. Each internal node keeps the information about the divisions performed in the node including the field used in the cutting, the number of cuttings, and the pointers to its children. Each leaf node keeps the rules relating to the space covered by

the node. In grid-of-tries structure, pointers are used instead of rule repetition to relate the nodes. This contributes to the reduction of memory usage. This method does not require recursive traversals; rather, it only traces the pointers back to the node. The algorithm's update time is so long that it is better to recreate the data structure from scratch for addition or omission of a rule. Therefore, this algorithm is appropriate for static packet classifiers in two dimensions, but it cannot be easily extended to multidimensional modes. An algorithm that is suitable for multidimensional modes is Cross-product. In this algorithm, for any given packet P, the best match for each header field is found and all of the results are finally combined to find the best match [27].

Another algorithm is Recursive Flow Classification. This algorithm works by mapping the packet header information onto a smaller number of bits in several phases according to the features of actual classifiers. It is suitable for large numbers of fields and provides a relatively high speed of access, but it has low scalability because it changes the structure of classification fields by adding a new field and requires hardware implementation which is usually difficult to modify [27].

C. Leaf Pushing

A leaf-pushed tree pushes all the prefixes in the internal nodes downward into the leaves, thus storing prefixes only in its leaves [29], [30].

None of the algorithms so far proposed have been able to compromise between classification time and memory usage. In other words, each of these algorithms is optimal either in terms of classification time or in terms of the memory used by its data structure. Therefore, we need a classification algorithm that would be efficient concerning both criteria. With this aim, the next section proposes such a method by making use of the best features of previous algorithms.

III. THE PROPOSED METHOD

In this section, first, we explain the basic KD-tree algorithm and its related data structure using a sample ruleset. Next, we explain how our proposed method applies the leaf-pushing technique on the sample KD-tree. Finally, a bloom-filter based implementation of the leaf-pushed KD-tree is completely explained.

A. KD-Tree Structure

In this algorithm each packet is represented as a point in the geometric space. Space decomposition algorithms provide a search technique that uses a tree structure to find a rule that covers the packet. In a tree structure, all children of a node share an identical prefix that is inherited from the parent node. For example, the children of a parent node that begins with "0" will begin with "0".

Fig. 1 shows an example of a space decomposed by the two fields F1 and F2 which represent the source and destination prefix addresses from Table I, respectively. The wild card state, represented by *, means that the rest of the bits can be 0 or 1.

TABLE I
EXAMPLE OF A RULE SET [30]

Rule No	Source Prefix	Destination Prefix	Source Port	Destination Port	Protocol Type
R0	010*	011*	0,65535	1704,1704	6
R1	01100*	0110*	161,161	1711,1711	6
R2	0110*	1001*	1024,1024	1521,1521	6
R3	1010*	1101*	119,119	1717,1717	6
R4	1*	10*	53,53	2110,2110	6
R5	00*	0*	1024,1024	1717,1717	6
R6	*	110*	80,80	1221,1221	6
R7	000*	*	0,65535	0,65535	6
R8	001*	00*	0,65535	0,65535	*
R9	00*	111*	0,65535	0,65535	*

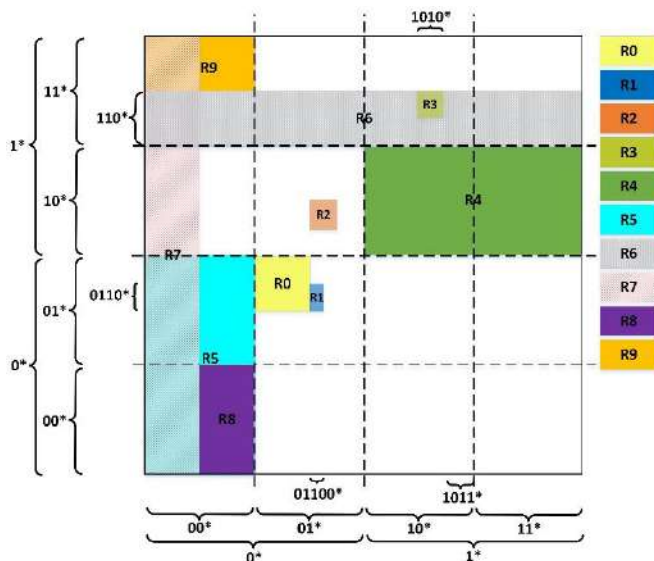


Fig. 1. The rules from Table I as represented in the geometric decomposition space of the KD-tree.

The space covered by a prefix on one axis is inversely related to the prefix length; that is, a shorter prefix covers a larger space. The length of the wild card state, for example, is always 0 and covers all the input spaces on the axis.

The partitioning of the geometric space is as following. The search space is recursively decomposed into two equal partitions based on F1 and F2. At the first level, this is done through F1 which is the first dimension and, at the second level, this is done through F2 which is the second dimension. Thus, if one of the corners of the square space of a rule crosses the boundary of its partition, the rule is considered as part of the Crossing Filter Set (CFS) of that partition.

A KD-tree makes combines recursive decomposition and tree-like structure. In fact, it provides two-dimensional packet classification for binary trees with the aim of searching an IP address.

As shown in Fig. 2, a KD-tree is built by the source and destination prefix address of a rule. Each level of the tree in the search space is divided into two parts based on one of the prefixes.

We begin by the root node which covers the entire search space. Partitioning at this level is based on the source prefix

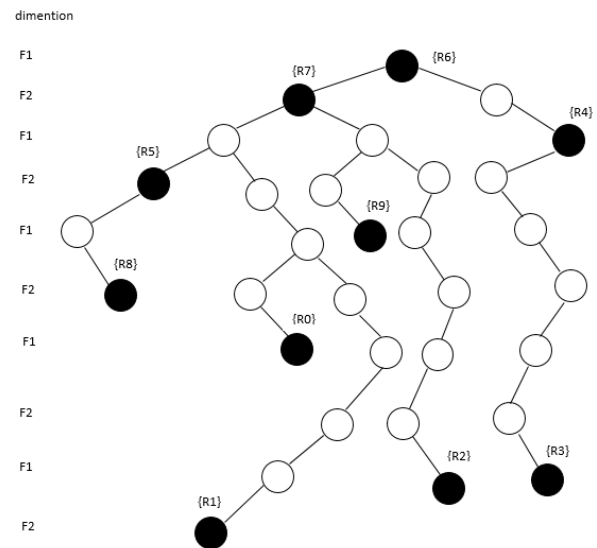


Fig. 2. The binary KD-tree of the geometric space represented in Fig. 1.

code. In this way, all the rules with a source prefix code of 0 (in the left-hand partition of the geometric space) are inserted on the left side of the root and the rules with a source prefix code of 1 (in the right-hand partition of the geometric space) are inserted on the right side. At the next level, partitioning is done on the basis of the destination prefix address. This will continue until every rule has been placed in a node. The inserted into the CFS of a partition have identical prefixes which are derived from the shortest prefix of each rule. They are inserted into a node where the area and path correspond to the sum of the lengths of source and destination prefix addresses and the value of the source and destination prefix codes, respectively. In this method, the rules are stored once without any repetition.

Note that the shortest length of two prefixes determines the area in which the rule is stored. In other words, the KD-tree does not exactly represent the decomposed space. This will increase the number of nodes on each path from the root to a leaf and decrease the efficiency of search. The reason is that the code used in the generation of a KD-tree is produced based on the length of the shortest prefix field of the rule and the rest of the length of longer fields is not used.

B. Leaf-Pushed KD-Tree

A leaf-pushed tree pushes all the prefixes in the internal nodes downward into the leaves. Therefore, prefixes are only stored in the leaves Fig. 3 represents the implementation of the leaf-pushing on the tree of Fig. 2. The prefixes in a leaf-pushed tree are joined, which optimizes the IP address search. Each leaf node in the leaf-pushed tree corresponds to the joined range of coverage and stores the prefixes which the range covers. The leaf-pushing technique used here differs from that utilized in IP address search problems. In IP address search where the longest prefix matching is at stake, only the longest prefixes are pushed into the leaves while here we push all prefixes that cover the same range into the leaves with the aim of solving packet classification problems.

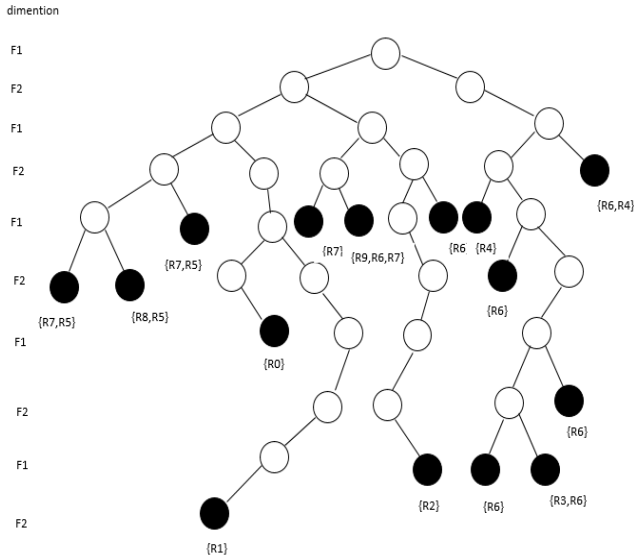


Fig. 3. The leaf-pushed tree of the geometric space represented in Fig. 1.

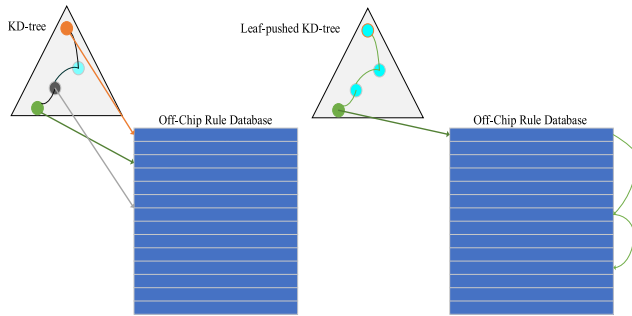


Fig. 4. Comparison of the architecture of conventional KD-tree and leaf-pushing tree.

In what follows, we seek to turn a KD-tree into a leaf-pushed tree. The leaf-pushed tree is created as following. In the example in Fig. 2, the rules stored in the internal nodes include R4, R5, R6, and R7. Let us examine the leaf-pushing process for R4 ($1^*, 10^*$). This rule is in the first dimension. Since there is no other prefix in this dimension, the rule can cover both the left and the right child. Therefore, if we extend the prefix address of the first dimension, which is the starting point, we will obtain 10^* and 11^* and the rule R4 will be transferred to its two child nodes. As the right node is a leaf, further extension on this side is not necessary. On the left side, as R4 still lies in an internal node, it should be further extended. Since this rule still has a prefix code on the second dimension (i.e. the destination), this bit will be used. The bit is 0. Therefore, we move to the left side of the node and stop further extension on arriving at a leaf node. This process will continue for all rules in the internal nodes. In fact, further extension of rules should stop with the end of their nested relations because, although further extension will increase search efficiency, the required memory will also increase due to the repetition of rules in the nodes.

Algorithm 1 shows the pseudo code for searching the leaf-pushed KD-tree. The input to this function is the input that was assumed for explaining the search process in this tree, i.e.

Algorithm 1 The Pseudo Code for Searching the Leaf-Pushed KD-Tree

```

Input: packet in_pkt
Output: rule R
1: function SearchLeafPushingKdtree(in_pkt)
2:   BMR = default
3:   next_node = root; i = 0
4:   while (next_node != NULL) do
5:     node = next_node
6:     if ((node.type =
           = RuleNode)&&(BMR
           > node.pri)) then
7:       BMR = linearSearch(in_pkt)
8:       break
9:     else
10:      if (node.dimension == 0) then
11:        | next_node = node.ptr(in_pkt.srcA[i])
12:      else if (node.dimension == 1) then
13:        | next_node = node.ptr(in_pkt.dstA[i])
14:        | i ++
15:      end if
16:    if end
17:  end while
18:  //search for wildcard rules
19:  if (BMR > wild.threshold) then
20:    | BMR = linearSearch(in_pkt);
21:  if end
22:  return BMR
23: function end

```

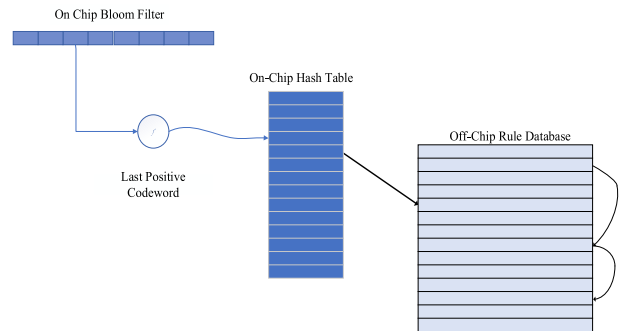


Fig. 5. Searching a tree by means of a Bloom filter and a hash table.

(6, 1711, 161, 01100, 01100). The output of the function is the best matching rule (BMR). First, a default value is determined for cases where the packet does not match any of the rules in the database (line 2). The default value is usually the wild-card state. Lines 4 to 17 traverse the tree. The traversal begins at the root and the first dimension. The algorithm takes the first bit of the source prefix code, which is 0, and moves to the left child (line 10). At the next level, it takes the first bit of the destination prefix code and moves to the left child (line 12). Then it takes the second bit of the source prefix code, which is 1, and moves to the right child. As the process continues and a leaf node containing R1 is achieved (line 6), the search is finished. R1 is compared with the packet header and, if they

TABLE II
COMPARISON OF THE BEHAVIOR OF CONVENTIONAL KD-TREE AND THE PROPOSED LEAF-PUSHING TREE

Size \ Rule set Type	KD-tree				Leaf-pushed KD-tree			
	5K	10K	50K	100K	5K	10K	50K	100K
ACL	4834	9835	49220	97450	4834	9835	49220	97450
	4834	9835	49220	97450	36838	46649	79361	145263
	14842	21789	8616	9107	25825	37003	10619	10921
IPC	4731	9533	32111	57104	4731	9533	32211	57104
	4731	9533	32211	57104	26265	62125	393463	655118
	22970	46349	4146	4399	40839	82333	4869	5011
FW	4710	9387	32578	44828	4710	9387	32578	44828
	4710	9387	32578	44828	194160	364919	439512	881245
	20052	40476	11742	1454	3143	7013	20571	1745

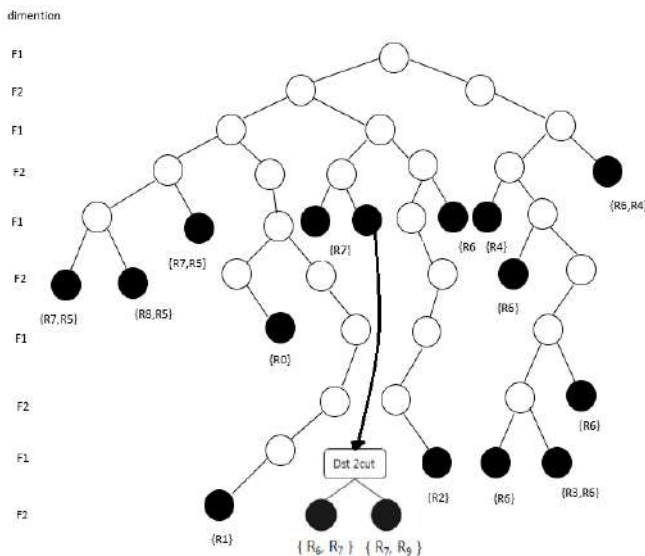


Fig. 6. The structure of the leaf-pushing KD-tree in Figure (3) as modified by means of HiCuts.

match, it is returned as the best matching rule (line 7). In this step, if there are several rules in the node, they are searched linearly to find the best matching rule. In this example, the search is finished only by comparing one rule. Comparison with R6 and R7 is avoided because they lie in the leaves. Lines 18 to 22 are executed when none of the rules in the tree match the packet. In this case, the packet is matched linearly against a list of rules in which both input fields have wild-card values and which have already been ordered by priority.

Fig. 4 compares the architecture of conventional KD-tree and leaf-pushed tree. It should be noted that we keep the

KD-tree in the on-chip memory and the database in the off-chip memory due to its large size. When a node containing a rule is observed in a KD-tree, the algorithm is referred to the memory whereas, in a leaf-pushed tree, the entire search process is performed within the on-chip memory. The pointer obtained in this search is used to access the off-chip memory which keeps the classifier’s database.

C. Generating a Leaf-Pushed KD-Tree by Using a Bloom Filter

In this section, we shall introduce a useful method for implementing a leaf-pushing KD-tree. Characteristic of this tree is that all the nodes that contain rules lie at the last level. Thus, an efficient search method is to use a Bloom filter and a hash table. Fig. 5 illustrates the proposed method which makes use of a Bloom filter, a hash table, and a rule database.

The Bloom filter is responsible for determining whether or not each input substring has a corresponding node in the tree. Therefore, the Bloom filter should be applied to all the nodes that contain rules in a leaf-pushing KD-tree.

First, the length of prefixes in the tree is sorted in a descending order and represented using vectors. Then a substring with the same length as the longest prefix in the tree is retrieved from the source and destination address prefixes of the packet and a query is sent to the Bloom filter. If the result is positive, the node with this prefix length contains a rule that matches the input. As a Bloom filter never produces false negatives, a negative result means that there is no node with the current length. Afterwards, further queries will be sent to the Bloom filter as the length of the input substring is being reduced down to smaller lengths in the prefix vector. This will continue until

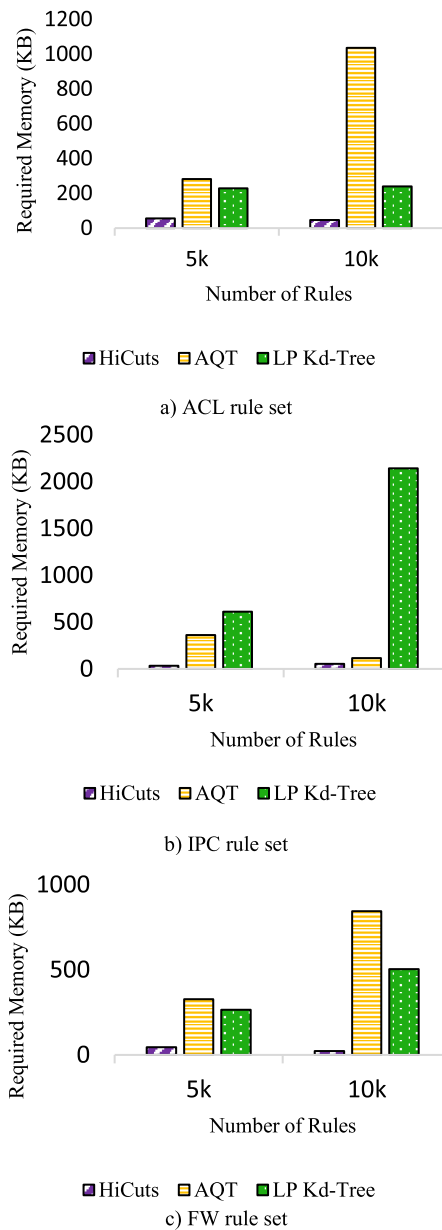


Fig. 7. Comparison of memory access among the proposed algorithm, HiCuts, and AQT.

a positive result is obtained. In this way, the search proceeds only by querying the Bloom filter. The role of the hash table is to provide a pointer to possibly matching rules in the database. For this purpose, every rule node must be stored in the hash table.

For example, let us assume the input packet (01100, 01100, 161, 1711, 6). In the tree in Fig. 3, the vector of prefix lengths is $\langle 3, 4, 5, 6, 7, 8, \text{ and } 9 \rangle$. The pseudo code for Bloom filter search is shown in Algorithm 2. The input to this function is our example packet. The output of the function is the best matching rule (BMR). The Bloom filter programmed according to the nodes of the tree in Fig. 3 will return a positive result (line 3 in Algorithm 2) for the substring 001111000*. Suppose that the probability of false positive results is sufficiently small. Using the substring 001111000

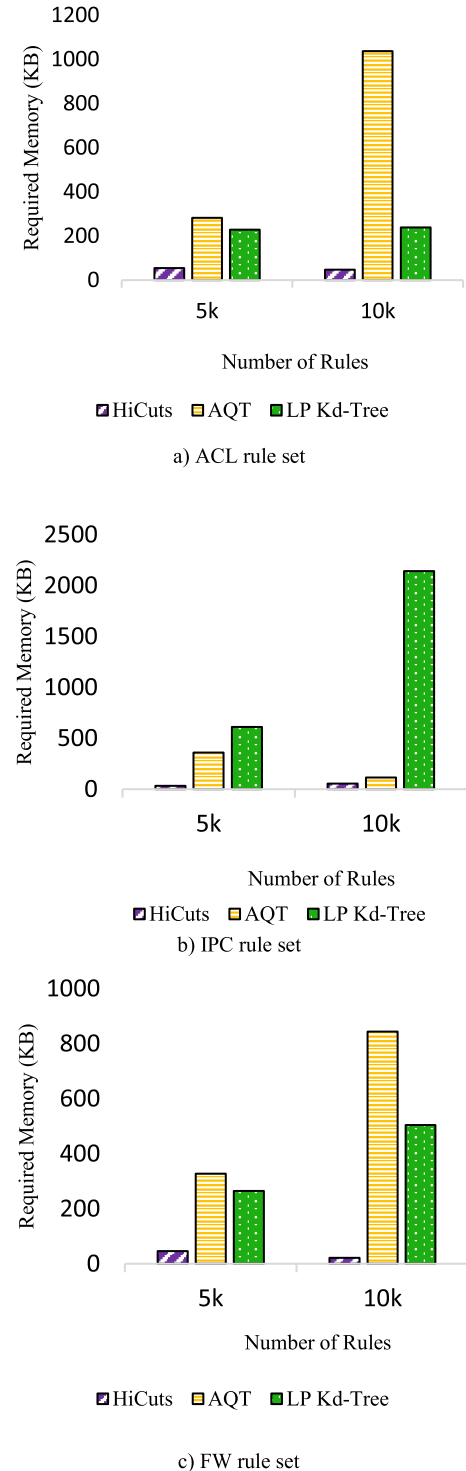


Fig. 8. Comparison of memory usage among the proposed algorithm, HiCuts, and AQT.

(which is a positive substring) as a hash key, the hash table is accessed (line 4). By obtaining a pointer from the hash table, R1 is accessed. Next, R1 is compared with the packet header and, if they match, it is returned as the best matching rule (line 6).

D. Modification of the Algorithm

Space decomposition algorithms such as HyperCuts [28], HiCuts [27], and BC [31] are controlled by a predetermined

TABLE III
COMPARISON OF THE MEMORY USED BY KD-TREE AND THE PROPOSED LEAF-PUSHED KD-TREE

Dataset	# Rules		KD-tree		Leaf-pushed KD-tree			
			On-chip	Off-chip	On-chip		Off-chip	
			M_i (KB)	M_r (MB)	M_b (KB)	M_h (KB)	Total (KB)	M_r (MB)
ACL	5K	4834	76	0.09	64	143	207	3.4
	10K	9835	119	0.185	128	186	314	5.42
	50K	49220	47	0.938	32	106	138	1.52
	100K	97450	51.13	1.87	32	111	142	2.08
IPC	5K	4731	123.37	0.08	128	220	348	4.28
	10K	9533	265.34	0.179	256	320	576	5.365
	50K	32211	21.25	0.61	16	50	66	7.64
	100K	57104	23.9	1.089	16	52	67	12.8
FW	5K	4710	93	0.088	128	392	519	2.9
	10K	9387	232.22	0.176	256	811	1066	3.96
	50K	32578	63	0.617	64	210	274	8.54
	100K	44828	6.8	0.855	4	18	22	8.54

Algorithm 2 The Pseudo Code for Searching by the Proposed Algorithm Using a Bloom Filter

Input: *packet in_pkt*

Output: *rules R*

```

1: function SearchWithBF (n_pkt)
2:   BMR = default
3:   BML = SearchBF (n_pkt)
4:   rulePtr = SearchHASH (BML)
5:   //rule database search
6:   BMR = linearSearch(in_pkt, rulePtr)
7:   //search for wildcard rules
8:   if (BMR > wild.threshold) then
9:     | BMR = linearSearch(in_pkt);
10:  end if
11:  return BMR
12: end function

```

proposed in the previous section is presented here so as to reduce the number of memory accesses. In this modified structure, a HiCuts tree is produced for each leaf node in which the number of rules is greater than the value of *binth*.

In other words, the space covered by each node in the leaf-pushing KD-tree is partitioned to make the number of rules in a decomposition space equal to or smaller than the value of *binth*. Fig. 6 represents the modified structure of the leaf-pushing KD-tree in Fig. 3, with *binth* set to 2. The space separated by the node 0101* contain three rules, which is greater than *binth*. As a result, this space is partitioned by HiCuts.

IV. IMPLEMENTATION AND EVALUATION

The proposed algorithm was implemented using C++ and Classbench Suite [32]. Two of the most important criteria used in the evaluation of packet classification algorithms include search time (which is directly related to the number of memory accesses) and memory usage. In our discussion, N denotes the number of rules in the database and W denotes the maximum prefix length in the rule database. Every rule has d dimensions.

A. Classbench

Classbench [33] is a simulator for generating rule sets with any distribution along with headers corresponding to the rules. This software suite can also produce the required packets. It performs this task by using the control information and the input parameters called ‘seed’ which are given to it through a text file. This simulator fulfills the need of the developers of packet classification algorithms for authentic, heterogeneous rules that are found in firewalls, IP chains, and Access Control Lists. In this study, we used three filter sets corresponding to the parameters *Acl2*, *Fw2*, and *Ipc2* with the number of rules being 5k, 10k, 50k, and 100k.

B. Metrics

In this section, the efficiency of the suggested algorithm is studied from different aspects such as memory required for storing the data structure, complexity of algorithm, and

TABLE IV
COMPARISON OF THE PROPOSED ALGORITHM WITH OTHER ALGORITHMS

Evaluation algorithms	Lookup time	Memory usage	Dimension scalability
Linear search [34]	$O(N)$	$O(N)$	unlimited
Grid-of-tries [34]	$O(W)$	$O(NW)$	2
Cross-producting [35]	$O(dW)$	$O(N^d)$	unlimited
Bit-parallelism [36]	$O(W \log N)$	$O(NW)$	2
Area-based Quad Tree [26]	$O(W)$	$O(NW)$	2
Fat-Inverted Segment [37]	$O((L + 1)W)$	$O(LN^{(1+1/L)})$	2
Segment tree [38]	$O(\log N)$	$O(N * \log N)$	2
RFC [39]	$O(d)$	$O(N^d)$	Unlimited
HiCuts [40]	$O(d)$	$O(N^d)$	Unlimited
Linear search on tuple [41]	$O(W^d)$	$O(N)$	Unlimited
Rectangle search [41]	$O(W)$	$O(NW)$	2
Binary search [34]	$O(\log^2 W)$	$O(N \log^2 W)$	2
Extended Grid-of-Tries [42]	$O(W)$	$O(NW)$	Unlimited
Leaf-pushed KD-tree	$O(d \log W)$	$O(Nd \log W)$	Unlimited

variable called *binth*. The *binth* controls the number of rules in a decomposed space. To provide a similar control mechanism, a modified form of the leaf-pushing KD-tree which was

TABLE V
COMPARISON OF THE SEARCH EFFICIENCY OF KD-TREE AND LEAF-PUSHING KD-TREE IN
TERMS OF THE NUMBER OF MEMORY ACCESSES (A: AVERAGE, W: WORST -CASE)

Dataset	# Rules	# packets	KD-tree				Leaf-pushed KD-tree			
			On-chip		Off-chip		On-chip		Off-chip	
			A_i	W_i	A_r	W_r	A_i	W_i	A_r	W_r
ACL	5K	13980	24	64	74	84	26.3	32	3.5	53
	10K	29205	27	64	164	240	26.8	33	3.9	4
	50K	48420	24	64	365	654	32.1	32	23	208
	100K	95340	26	64	594	742	27	33	124	214
IPC	5K	20610	25	64	182	224	25	31	11	34
	10K	24336	25	64	142	394	25	32	14	30
	50K	49047	28	64	524	642	28	33	43	245
	100K	94370	17	64	348	874	17	32	64	324
FW	5K	9201	7	64	589	784	7	30	89	98
	10K	13054	7	64	320	1247	7	31	20	125
	50K	49163	8	64	2312	5864	8	33	312	438
	100K	82473	9	64	3252	9865	9	30	352	569

maximum number of memory accesses in classifying a typical packet.

C. Evaluation

Table II compares the behavior of conventional KD-tree and leaf-pushing tree. While the number of nodes in the leaf-pushing tree does not show remarkable increase, the number of stored rules has significantly increased. The efficiency of the leaf-pushing tree strongly depends on the type of classifier as well as on the number of rules in the wild-card field because these rules tend to appear in many leaves. The FW rule set has a high rate of rule repetition.

Table III shows the memory required by the KD-tree and the leaf-pushing KD-tree. The size of the required on-chip memory (M_i) is calculated based on the width of a single node which includes the node type field, two child pointers, and one rule pointer. This width is then multiplied by the number of the nodes in the tree. This size is measured in KB, as opposed to the size of the off-chip memory which is measured in MB. As can be observed in the table, the increase in memory size is quite remarkable and the generated tree can be easily stored in the on-chip memory.

Table IV compares the complexity of the proposed algorithm with state-of-the-art algorithms. The total number of tuples in the classifier is Wd . The height of the KD-tree is $\log Wd$ or $d \log W$. The complexity of the structure is equal to the height of the balanced KD-tree, i.e. $O(d \log W)$. For the storage of N filters, the space complexity is $O(Nd \log W)$. Also, Table IV provides a comparison between the proposed algorithm and other classification algorithms. The proposed algorithm has an acceptable performance in terms of time and space complexity.

The average number of queries is related to the tree depth. The number of inputs to each rule set is shown in Table V.

The average number of rule comparisons is obtained by dividing the sum of comparisons for all inputs by the total number of inputs. The worst case of rule comparisons belongs to the input that causes the highest number of comparisons. Our evaluations show that the average number of access to the hash table in our algorithm is 1. The worst case of access

to the hash table is the maximum number of back-tracking as a result of the false positive of the Bloom filter. In this table, the number of accesses to the Bloom filter and the hash table is represented by A_i and W_i , respectively. The number of rule comparisons strongly depends on the type of sets and the features of the tree, particularly in the case of rules in which both prefix fields have the wild-card parameter. For example, the FW rule set has many such rules. As these rules are matched against the inputs after the BMR has been obtained from the leaf-pushing tree, the worst number of rule comparisons can be greater than the maximum number of rules in a leaf node. According to the results of our evaluations, the speedup achieved by the leaf-pushing KD-tree was 1 to 42 times as large as that achieved by the KD-tree. Fig. 7 compares the average number of accesses to the memory in each algorithm which refers to the number of rule comparisons. As can be seen in the figure, the proposed algorithm had a better performance in most of the sets in comparison with other algorithms. The reason is that in a Bloom filter with a remarkably low amount of error, access to the hash table is minimized. Moreover, since the numbers are sorted by their priority in the rule set, the number of rule comparisons is reduced as a result of decreased memory access. It can be seen in the figure that the number of memory accesses in the AQT algorithm has been reduced from 23 to 1.

In Fig. 8, the memory usage of the proposed modified structure is compared with that of HiCuts and AQT. Memory usage is directly related to the repetition of rules. The proposed modified structure can also be stored in an on-chip memory. Even if the on-chip memory is not sufficient, the significant reduction in the number of rule comparisons makes it possible to store the rule database in an off-chip memory without any concern about decrease in efficiency. As mentioned earlier, the number of stored rules has increased in the proposed method. The efficiency of the leaf-pushing tree strongly depends on the type of classifier as well as on the number of rules in the wild-card field because these rules tend to appear in many leaves. As the rate of rule repetition in FW and IPC rule sets is high, the memory usage of the proposed algorithm increases in these classifiers. As can be seen in the figure, the

memory usage of the algorithm is acceptable and there is a 77-percent reduction in comparison with AQT.

V. CONCLUSION

Software-defined intelligent vehicular networks require fast packet classification algorithms to provide several flow-based surveillance services to mobile applications on vehicular nodes. This requirement emerges when the scale of such networks grows exponentially and consequently results in a considerable delay in processing big streams of network packets to/from vehicular nodes. Using appropriate packet classification methods and enhancing their speed is a key solution to this problem.

In this paper, we first described the implementation of KD-tree which is an algorithm for packet classification and then discussed the structure of leaf-pushing tree. By using leaf-pushing, the prefix information in longer prefixes would significantly reduce the number of rules in a search path. The rules are kept only in leaf nodes. We showed that leaf-pushing technique can be efficiently used to separate the search process from the process of rule matching. To improve the performance of a previously generated tree, we used a Bloom filter and a hash table. The Bloom filter is used in our proposed method to search for a node that contains a rule that matches an incoming packet. The function of the hash table is to provide a pointer to the rule database when a node has been found to contain a matching rule. Finally, we also proposed a modified structure for our leaf-pushing KD-tree to enhance its performance and reduce the number of accesses to the off-chip memory.

We evaluated our method in terms of memory usage and memory access. Although the required memory increased only slightly, a significant improvement was observed in memory access. The obtained speedup is indicative of the efficiency of the proposed method. We compared the implementation results with other algorithms for geometric space decomposition such as AQT and HiCuts. The comparison proves that our modified structure is significantly more efficient in reducing the number of memory accesses. Our method could reduce this number from 23 to 1 and its memory usage was comparable to other algorithms.

To continue this research, parallel platforms like GPUs can be used for parallelization of the packet classification process. Given the larger number of computational cores in GPUs, it is predictable that the parallelization of the proposed algorithm would be expressively optimized on GPUs.

REFERENCES

- [1] M. Sredynski, G. Arnould, and D. Khadraoui, "The emerging applications of intelligent vehicular networks for traffic efficiency," in *Proc. 3rd ACM Int. Symp. Design Anal. Intell. Veh. Netw. Appl. (DIVANet)*, 2013, pp. 101–108.
- [2] M. B. Younes and A. Boukerche, "Safety and efficiency control protocol for highways using intelligent vehicular networks," *Comput. Netw.*, vol. 152, pp. 1–11, Apr. 2019.
- [3] J. Cheng, J. Cheng, M. Zhou, F. Liu, S. Gao, and C. Liu, "Routing in Internet of vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2339–2352, Oct. 2015.
- [4] Z. Zhou, X. Chen, Y. Zhang, and S. Mumtaz, "Blockchain-empowered secure spectrum sharing for 5G heterogeneous networks," *IEEE Netw.*, vol. 34, no. 1, pp. 24–31, Jan. 2020.
- [5] W. Xu, H. Zhou, H. Wu, F. Lyu, N. Cheng, and X. Shen, "Intelligent link adaptation in 802.11 vehicular networks: Challenges and solutions," *IEEE Commun. Standards Mag.*, vol. 3, no. 1, pp. 12–18, Mar. 2019.
- [6] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6G: Machine-learning approaches," *Proc. IEEE*, vol. 108, no. 2, pp. 292–307, Feb. 2020.
- [7] X. Lin, J. Wu, S. Mumtaz, S. Garg, J. Li, and M. Guizani, "Blockchain-based on-demand computing resource trading in IoV-assisted smart city," *IEEE Trans. Emerg. Topics Comput.*, early access, Feb. 6, 2020, doi: 10.1109/TETC.2020.2971831.
- [8] G. Raja, A. Ganapathisubramanian, S. Anbalagan, S. B. M. Baskaran, K. Raja, and A. K. Bashir, "Information-driven reward-based data offloading in next-generation vehicular networks," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3747–3758, May 2020.
- [9] L. Nkenyereye, L. Nkenyereye, S. M. R. Islam, Y.-H. Choi, M. Bilal, and J.-W. Jang, "Software-defined network-based vehicular networks: A position paper on their modeling and implementation," *Sensors*, vol. 19, no. 17, p. 3788, Aug. 2019.
- [10] F. A. Silva, A. Boukerche, T. R. M. B. Silva, E. Cerqueira, L. B. Ruiz, and A. A. F. Loureiro, "Information-driven software-defined vehicular networks: Adapting flexible architecture to various scenarios," *IEEE Veh. Technol. Mag.*, vol. 14, no. 1, pp. 98–107, Mar. 2019.
- [11] J. Bhatia, Y. Modi, S. Tanwar, and M. Bhavsar, "Software defined vehicular networks: A comprehensive review," *Int. J. Commun. Syst.*, vol. 32, no. 12, p. e4005, Aug. 2019.
- [12] J. C. Nobre *et al.*, "Vehicular software-defined networking and fog computing: Integration and design principles," *Ad Hoc Netw.*, vol. 82, pp. 172–181, Jan. 2019.
- [13] T. Ganegedara and V. K. Prasanna, "StrideBV: Single chip 400G+ packet classification," in *Proc. IEEE 13th Int. Conf. High Perform. Switching Routing (HPSR)*, Jun. 2012, pp. 1–6.
- [14] P. Gupta and N. McKeown, "Algorithms for packet classification," *IEEE Netw.*, vol. 15, no. 2, pp. 24–32, Mar./Apr. 2001.
- [15] C.-L. Hsieh and N. Weng, "Scalable many-field packet classification using multidimensional-cutting via selective bit-concatenation," in *Proc. 11th ACM/IEEE Symp. Archit. Netw. Commun. Syst. (ANCS)*, May 2015, pp. 187–188.
- [16] W. Jiang and V. K. Prasanna, "Scalable packet classification on FPGA," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 9, pp. 1668–1680, Sep. 2012.
- [17] Y. R. Qu, H. H. Zhang, S. Zhou, and V. K. Prasanna, "Optimizing many-field packet classification on FPGA, multi-core general purpose processor, and GPU," in *Proc. ACM/IEEE Symp. Archit. Netw. Commun. Syst. (ANCS)*, May 2015, pp. 87–98.
- [18] B. S. Tumari and W. LakshmiPriya, "FPGA implementation of binary-tree-based high speed packet classification system," *Int. J. Combined Res. Develop.*, vol. 2, no. 6, pp. 17–22, Jun. 2014.
- [19] K. Zheng, H. Che, Z. Wang, and B. Liu, "TCAM-based distributed parallel packet classification algorithm with range-matching solution," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, Mar. 2005, pp. 293–303.
- [20] K. Zheng, H. Che, Z. Wang, B. Liu, and X. Zhang, "DPPC-RE: TCAM-based distributed parallel packet classification with range encoding," *IEEE Trans. Comput.*, vol. 55, no. 8, pp. 947–961, Aug. 2006.
- [21] Z. Cao, M. Kodialam, and T. V. Lakshman, "Traffic steering in software defined networks: Planning and online routing," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 65–70, Feb. 2015.
- [22] K. G. Perez, X. Yang, S. Scott-Hayward, and S. Sezer, "A configurable packet classification architecture for software-defined networking," in *Proc. 27th IEEE Int. Syst.-Chip Conf. (SOCC)*, Sep. 2014, pp. 353–358.
- [23] S. Han, K. Jang, K. Park, and S. Moon, "PacketShader: A GPU-accelerated software router," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 195–206, 2011.
- [24] K. G. Perez, X. Yang, S. Scott-Hayward, and S. Sezer, "Optimized packet classification for software-defined networking," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 859–864.
- [25] H. Lim, Y. Choe, M. Shim, and J. Lee, "A quad-trie conditionally merged with a decision tree for packet classification," *IEEE Commun. Lett.*, vol. 18, no. 4, pp. 676–679, Apr. 2014.
- [26] H. Lim, M. Y. Kang, and C. Yim, "Two-dimensional packet classification algorithm using a quad-tree," *Comput. Commun.*, vol. 30, no. 6, pp. 1396–1405, Mar. 2007.
- [27] D. Pao and Z. Lu, "A multi-pipeline architecture for high-speed packet classification," *Comput. Commun.*, vol. 54, pp. 84–96, Dec. 2014.

- [28] S. Singh, F. Baboescu, G. Varghese, and J. Wang, "Packet classification using multidimensional cutting," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun.*, 2003, pp. 213–224.
- [29] V. Srinivasan and G. Varghese, "Fast address lookups using controlled prefix expansion," *ACM Trans. Comput. Syst.*, vol. 17, no. 1, pp. 1–40, Feb. 1999.
- [30] J. Lee, H. Byun, J. H. Mun, and H. Lim, "Utilizing 2-D leaf-pushing for packet classification," *Comput. Commun.*, vol. 103, pp. 116–129, May 2017.
- [31] H. Lim, N. Lee, G. Jin, J. Lee, Y. Choi, and C. Yim, "Boundary cutting for packet classification," *IEEE/ACM Trans. Netw.*, vol. 22, no. 2, pp. 443–456, Apr. 2014.
- [32] D. E. Taylor and J. S. Turner, "ClassBench: A packet classification benchmark," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 499–511, Jun. 2007.
- [33] D. E. Taylor and J. S. Turner, "ClassBench: A packet classification benchmark," in *Proc. 24th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, Mar. 2005, pp. 2068–2079.
- [34] D. E. Taylor, "Survey and taxonomy of packet classification techniques," *ACM Comput. Surv.*, vol. 37, no. 3, pp. 238–275, Sep. 2005.
- [35] V. Srinivasan, G. Varghese, S. Suri, and M. Waldvogel, "Fast and scalable layer four switching," in *Proc. ACM SIGCOMM Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM)*, 1998, pp. 191–202.
- [36] T. V. Lakshman and D. Stiliadis, "High-speed policy-based packet forwarding using efficient multi-dimensional range matching," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 28, no. 4, pp. 203–214, Oct. 1998.
- [37] A. Feldman and S. Muthukrishnan, "Tradeoffs for packet classification," in *Proc. IEEE Conf. Comput. Commun., 19th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, Mar. 2000, pp. 1193–1202.
- [38] C.-F. Su, "High-speed packet classification using segment tree," in *Proc. IEEE Global Telecommun. Conf. (Globecom)*, vol. 1, Nov./Dec. 2000, pp. 582–586.
- [39] P. Gupta and N. McKeown, "Packet classification on multiple fields," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM)*, 1999, pp. 147–160.
- [40] P. Gupta and N. McKeown, "Packet classification using hierarchical intelligent cuttings," in *Proc. Hot Interconnects*, 1999, pp. 1–9.
- [41] V. Srinivasan, S. Suri, and G. Varghese, "Packet classification using tuple space search," in *Proc. Conf. Appl., Technol., Archit., Protocols Comput. Commun. (SIGCOMM)*, 1999, pp. 135–146.
- [42] H. Lu and S. Sahni, " $O(\log W)$ multidimensional packet classification," *IEEE/ACM Trans. Netw.*, vol. 15, no. 2, pp. 462–472, Apr. 2007.



Mahdi Abbasi received the B.Sc., M.Sc., and Ph.D. degrees in computer engineering from the Sharif University of Technology, Tehran, Iran, and the University of Isfahan, Isfahan, Iran, respectively. He is currently with the Department of Computer Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan, Iran. His research interests include computer architecture, signal and image processing, machine learning, the Internet of Things (IoT), and computer networks.



Hajar Rezaei received the M.Sc. degree in computer engineering (computer networks) from Bu-Ali Sina University, Hamedan, Iran, in 2019. Her research interests include computer networks, software defined networking, and the Internet of Things (IoT).



Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include sensors, the IoT, fog computing, and underwater acoustic sensor networks. He is also a Distinguished Speaker of ACM.



Lianyong Qi (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Technology, Nanjing University, China, in 2011. He is currently an Associate Professor with the School of Information Science and Engineering, Qufu Normal University, China. He has already published more than 50 articles, including JSAC, TCC, TBD, FGCS, JCSS, CCPE, and ICWS. His research interests include services computing, big data, and the IoT.



Mohammad R. Khosravi is currently with the Department of Computer Engineering, Persian Gulf University, Iran. His main interests include statistical signal and image processing, medical bioinformatics, radar imaging and satellite remote sensing, computer communications, industrial wireless sensor networks, underwater acoustic communications, information science, and scientometrics.

Efficient Flow Processing in 5G-Envisioned SDN-Based Internet of Vehicles Using GPUs

Mahdi Abbasi¹, Ali Najafi, Milad Rafiee², Mohammad R. Khosravi³,

Varun G. Menon⁴, *Senior Member, IEEE*, and Ghulam Muhammad⁵, *Senior Member, IEEE*

Abstract—In the 5G-envisioned Internet of vehicles (IoV), a significant volume of data is exchanged through networks between intelligent transport systems (ITS) and clouds or fogs. With the introduction of Software-Defined Networking (SDN), the problems mentioned above are resolved by high-speed flow-based processing of data in network systems. To classify flows of packets in the SDN network, high throughput packet classification systems are needed. Although software packet classifiers are cheaper and more flexible than hardware classifiers, they could only deliver limited performance. A key idea to resolve this problem is parallelizing packet classification on graphical processing units (GPUs). In this paper, we study parallel forms of Tuple Space Search and Pruned Tuple Space Search algorithms for the flow classification suitable for GPUs using CUDA (Compute Unified Device Architecture). The key idea behind the offered methodology is to transfer the stream of packets from host memory to the global memory of the CUDA device, then assigning each of them to a classifier thread. To evaluate the proposed method, the GPU-based versions of the algorithms were implemented on two different CUDA devices, and two different CPU-based implementations of the algorithms were used as references. Experimental results showed that GPU computing enhances the performance of Pruned Tuple Space Search remarkably more than Tuple Space Search. Moreover, results evinced the computational efficiency of the proposed method for parallelizing packet classification algorithms.

Index Terms—5G, flow processing, GPU, Internet of Vehicles, intelligent transport systems, SDN.

I. INTRODUCTION

OUR intelligent vehicular world is connected by the Internet of vehicles (IoV). With the advent of the fifth-generation (5G) wireless networks, the significant development of network bandwidth and growth of hardware

technologies has increased the speed of communications considerably [2]–[4]. That is, the communications speed is reached to terabits per second. SDN enhances the performance by accelerating network systems to process the packets with the rate of vehicular network communications [3], [5], [6]. For this purpose, a new technology, packet classification, is used in the architecture of modern network systems, which lets them be flow-aware. In such systems, after classifying the received packets into flows, the corresponding actions are performed on each flow. A variety of modern network systems, including high-speed core routers, network firewalls, intelligent intrusion detection systems, and high-level network management systems, run packet classification as their fundamental process. In SDN-based IoV, as shown by Fig. 1, the flows of data produced by billions of IoV sensing devices are transferred to SDN controllers via high-speed switches and routers [7]. The intermediate SDN switches should apply flow-based actions on the received streams and process them at the speed of vehicular network links. Therefore, to accelerate the flow classification as the fundamental process of these systems, GPUs with their rich computational resources are highly interested.

To classify an incoming packet, first, its header information is compared against the filters of the classifier according to specified fields to find a match. In this comparison, more than one filter may match the packet, or no match may be found. In the former case, the priority of filters is used to select the best-matched filter. The action of this filter is applied to the packet. In the latter case, a default action is applied by the network processor on the unclassified packet. The standard fields extracted from the packet header to be used in packet classification include Service Type, Destination IP, Source IP, Destination Port, and Source Port. The classifier may access to other fields of the packet header like Source MAC and Destination MAC addresses [8].

Packet classification algorithms are generally implementable in hardware and software. Heavy computations of packet classification are performed on parallel processors and mainly, graphics processing units (GPUs) [9]. GPUs, these new emerging commodities, have increased the computational speed of modern systems compared to multi-core processors. By the introduction of useful modules like CUDA (Compute Unified Device Architecture) [10] and the OpenCL (Open Computing Language) [11], many pieces of research have focused on using GPUs to solve computation-intensive problems in various domains of science.

Manuscript received March 22, 2020; revised August 11, 2020 and October 4, 2020; accepted October 30, 2020. Date of publication December 7, 2020; date of current version August 9, 2021. The Associate Editor for this article was S. Garg. (*Corresponding author: Mahdi Abbasi.*)

Mahdi Abbasi, Ali Najafi, and Milad Rafiee are with the Department of Computer Engineering, Engineering Faculty, Bu-Ali Sina University, Hamedan 65178-38695, Iran (e-mail: abbasi@basu.ac.ir; a.najafi92@basu.ac.ir; m.rafee@alumni.basu.ac.ir).

Mohammad R. Khosravi is with the Department of Computer Engineering, Persian Gulf University, Bushehr 75168, Iran, and also with the Telecommunications Group, Shiraz University of Technology, Shiraz 71557-13876, Iran (e-mail: m.r.khosravi.taut@gmail.com; mohammadkhosravi@acm.org).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: varunmenon@ieee.org).

Ghulam Muhammad is with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University (KSU), Riyadh 11451, Saudi Arabia (e-mail: ghulam@ksu.edu.sa).

Digital Object Identifier 10.1109/TITS.2020.3038250

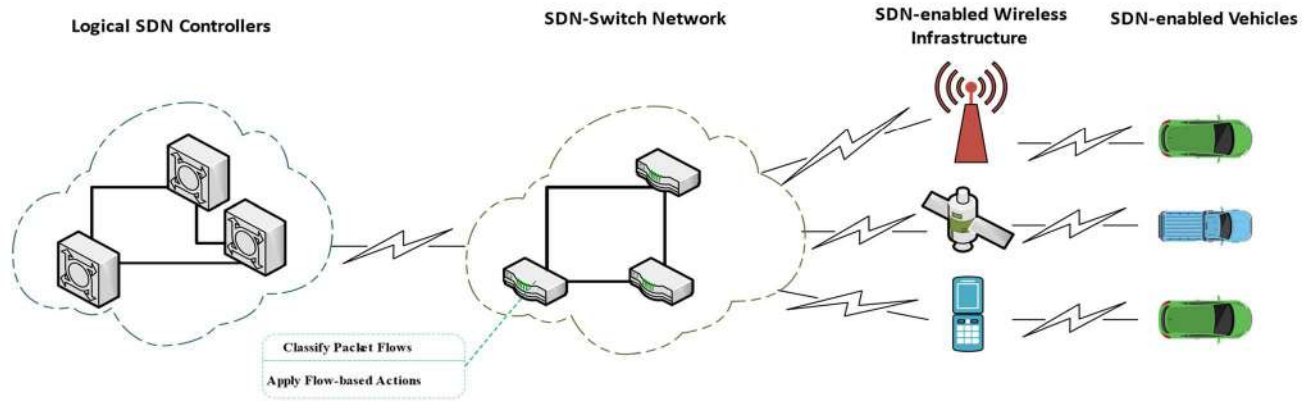


Fig. 1. SDN-Enabled internet of vehicles.

In this paper, we investigate GPUs to accelerate software-based packet classification algorithms. Two typical packet classification algorithms are selected for their simplicity and highly parallelizable structures. The current work has the following contributions:

- Different scenarios have been offered for the parallelization of two important algorithms of packet classification. The difference between the proposed plans is how the GPU uses different memory configurations. Due to the limited memory capacity and the data structure size of packet classification algorithms, a suitable method is proposed for optimal use of the GPU.
- The complexity of any parallelization scenario is highly dependent on using the memory hierarchy of GPU and the technique of exploiting the maximum concurrency among threads. Hence, we present an asymptotic analytic model to estimate the computational complexity of the proposed parallelization scenarios. This model helps us to compare the efficiency of different scenarios before running their corresponding kernels.

The rest of this paper is structured as follows. In the next section, related works on parallel packet classification via GPU are reviewed. In section 3, structures of TSS and TPS algorithms are evaluated for parallelization. In section 4, after establishing the CUDA programming model, the proposed method to parallelize TSS and TPS is explained. Experiments and their results are offered and analyzed in section 5. The conclusions and future works are presented in the final section.

II. RELATED WORK

Different classification algorithms, implementable in hardware or software, fall into one of the four categories, including *linear search*, *decision tree-based*, *decomposition-based*, and *tuple space-based* [8].

In all of the algorithms mentioned above, the classification process is launched per packet. That is, to classify each input packet, the classification algorithm is executed again. This key observation motivated limited researches to investigate the parallel forms of packet classification algorithms.

Nottingham and Irwin [11] provided valuable articles to pioneer the concept of parallel packet classification on GPU using CUDA and OpenCL platforms. However, their work does not include any implementation of algorithms and related performance comparisons. Han *et al.* [12] showed that by utilizing GPU co-processors, one might increase the classification throughput. Their proposed GPU-based IP routing algorithm, named PacketShader, could improve the performance compared with CPU-based IP routing methods. Hung *et al.* [13] presented parallelized versions of BPF and BitMap algorithms on GPUs. They also, utilized different memory architectures for GPU and compared numerical results revealing related computation performance. Kang and Deng [14] parallelized linear search and DBS algorithms and implemented them on GPU. Comparing the computation time of sequential versions of algorithms with that of GPU-based versions, they noticed that the performance of linear search on GPU is higher than DBS on GPU. Zhou *et al.* [9], [15] implemented the Bit-Vector algorithm on GPU. Their results showed that the performance of the algorithm, utilizing K computation threads on GPU, is enhanced to $\log_2 k$ times. Recently, Varvello *et al.* [16] used GPU computations to accelerate the Bloom filter algorithm. More recently, Zheng *et al.* [17] has presented an analogous study to enhance the performance of the HiCuts algorithm via parallelizing it on GPU.

Considering the researches mentioned above, some packet classification algorithms have not yet been considered for implementation on GPU. Moreover, none of the aforementioned studies have offered a general methodology to parallelize every packet classification algorithm. In the following, after reviewing the structure of the TSS and TPS, a general methodology is presented that simply makes the GPU-based implementation of these packet classification algorithms feasible.

III. BACKGROUND

A. Tuple Space Search

Srinivasan *et al.* [18] introduced the TSS technique for packet clarification. The key idea behind this method is to make the n the scope of a search on multiple fields of packet

TABLE I
TUPLE SPACE

Filter	(Source Prefix, Destination Prefix)	Tuple
R1	(00*,00*)	(2,2)
R2	(0*,01*)	(1,2)
R3	(1*,0*)	(1,1)

TABLE II
FILTER-SET

Filter	Src IP	Dst IP	Src Port	Dst Port	Protocol
R0	0001*	00001*	0,65535	25,25	6
R1	010*	0010*	53,53	443,443	4
R2	010*	001*	53,53	1024,65535	17
R3	110*	00*	53,53	443,443	4
R4	1111*	1*	53,53	25,25	4
R5	0101*	0*	0,65535	2788,2788	17
R6	0*	1101*	53,53	5632,5632	6
R7	*	10*	53,53	25,25	6
R8	1*	*	0,65535	2788,2788	17

header narrower by dividing the filters to mutually exclusive subsets according to definable “tuples”. Each tuple is a list of n values; each of them is the length of a field in a filter. For example, in filter-set on five prefix fields, the tuple [3, 5, 7, 0, 12] shows that the size of the first prefix field is 3 bits, the size of the second prefix is 5 bits, the size of the next prefix is 7, the size of the fourth one is 0 (or a wildcard field), and the size of the prefix of final field is 12.

TSS algorithm reduces the number of distinct tuples as compared with the number of filters in the original filter-set. Therefore, the filters of a filter-set are partitioned into the distinct tuple groups. Then, the algorithm performs lookups across all the identical tuples to find the most suitable filter that is highly matched with the packet. Finally, among the multiple reported filters, the highest priority one is reported as the best-matched filter.

In order to visualize the idea of tuples in the TSS algorithm, a sample filter-set is presented in Table I. The prefix of the source IP address and destination IP address of three filters are presented in Table I. The tuple of each filter is formed by putting the lengths of those prefixes together according to a predefined order. For example, since the size of the source and destination IP prefixes are 1 and 2, respectively, the tuple of R2 is (1,2).

In order to show the classification of packets with TSS, a sample filter-set containing nine filters is presented in Table II. Each filter of this table has five fields. Two binary trees are constructed using source and destination IP prefixes. For this purpose, according to the presence of 0 or 1 in the successive bits of the prefix, a branch is added to the left or right of the under-construction node of the tree.

Fig. 2 shows the binary trees corresponding to the source address field and the destination address field of the sample filter-set. Black nodes represent filters. Alongside these nodes, several relevant nodes are also provided.

The algorithm works for input (11010, 00001, 53, 443, 4) with a set of filters in Table II as follows. Input (11010, 00001, 53, 443, 4) contains five fields of IP headers. These fields are shown from left to right in Table II, respectively.

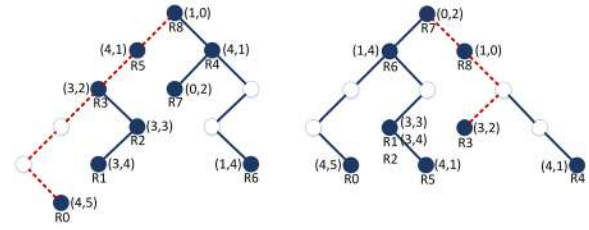


Fig. 2. The binary tree of source and destination addresses.

TABLE III
SELECTED FILTERS OF THE FILTER-SET

Filter	Src. IP	Dst. IP	Src. Port	Dst. Port	Protocol
R0	0001*	00001*	0,65535	25,25	6
R3	110*	00*	53,53	443,443	4
R4	1111*	1*	53,53	25,25	4
R5	0101*	0*	0,65535	2788,2788	17
R7	*	10*	53,53	25,25	6
R8	1*	*	0,65535	2788,2788	17

TABLE IV
SELECTED TUPLES FROM PRUNED TUPLE SPACE

Filter	Src.IP IP	Dst. IP	Src. Port	Dst. Port	Protocol
R3	110*	00*	53,53	443,443	4
R8	1*	*	0,65535	2788,2788	17

In this example, the search is performed by traversing the constructed binary trees using corresponding values extracted from the packet header. The dashed line in Fig.2 shows the search path. The filters corresponding to the tuples found in the navigation path are extracted.

Table III shows the extracted filters for this example. As can be seen, the number of these filters is a small fraction of the filter-set. As a result, a substantial depletion of the irrelevant filters in the large-size filter-sets is expectable.

Finally, to find the best-matched filter, a linear search is conducted on the small subset of filters that were resulted from the previous step. For the above example, the best adaptive filter is the R4 filter.

B. Pruned Tuple Space Search (TPS)

The original tuple space search algorithm performs a linear search on all tuples obtained from the lookups over two binary trees. But the pruned tuple space search algorithm first shares some of the results obtained by traverse in the source address and destination address tree. Then, an exhaustive search is performed on the small subset of filters that are found from the intersection of tuples. Table IV shows the result of applying the TPS algorithm to the example of Table III.

Comparison of Tables III and IV shows that the speed of the TPS algorithm is higher than that of the TSS algorithm. This is well illustrated in the evaluations carried out. But the speed of this algorithm is still far from the ideal speed. One of the best ways to accelerate the packet classification algorithms is to execute them in parallel on graphics processing units. In the last few years, there has been some work on parallelizing

packet sorting algorithms to the graphics processing unit, which will be reviewed later.

It can be easily deduced from the above discussion that the TSS and TPS algorithms for packet classification benefit significantly from the parallelizable search. In the following section, parallel implementation of the TSS and TPS packet classification algorithms is explained.

C. CUDA Programming Model

Generally, the GPUs are processing systems in which many scalar processors execute threads of code in parallel. Modern GPUs include a number of stream processors (SMs) that manage some scalar processors (SPs). To utilize the computational power of these processors, standard platforms like CUDA are provided by Nvidia [10], [19].

CUDA programs classically include two related modules, a commonly sequential module that is executable on the CPU of the system, or the host, and a heavier-but-parallel module called the kernel that runs on the SPs. The former controls the transfer of the input data for the latter from system memory to the GPU memory hierarchy. Also, it copies the computation results from the memory hierarchy of the device to the system memory. The programmer should concisely allocate the required kernel memory and minimize the kernel communication overhead.

Kernel invokes a predefined set of threads on SPs that compute results for a slice of input data. To handle these threads, several blocks of thread are defined, each of which contains 512 concurrent threads. The blocks of threads are typically organized in a two/three-dimensional grid. By using the unique index of each thread, the data elements which should be processed by that thread are specified. Each block of threads is run by a single multiprocessor, which orchestrates the execution of the kernel on the corresponding SPs and coordinates the access of threads to input data elements and also storing the computation results through shared memory.

CUDA devices have different memory modules with varying delays of access. The performance of the kernel is positively related to the memory modules used in the kernel code.

Fig. 3 illustrates the proposed plan to keep filter-sets and packets in the CUD device. Given the size of the filter-set and packet headers, we prefer to store the filter-set and packets in the shared memory and the global memory of the device, respectively. The CUDA application programming interface (API) coordinates the memory access on the device, and provides a set of functions to communicate the data between the host memory and allocated device memory. Note that, among different data-transfer models, the streaming model is the best. It is a pipeline of data transmission. In this model, sequences of operations are scheduled to be performed progressively on the CUDA device [10], [20]. This model is used for data transfer in the proposed parallelization methodology.

IV. PARALLEL IMPLEMENTATION OF TSS AND TPS

This paper presents three different methods for parallelizing the search using a GPU. The first method uses global memory, and the two other methods use shared memory. Our GPU

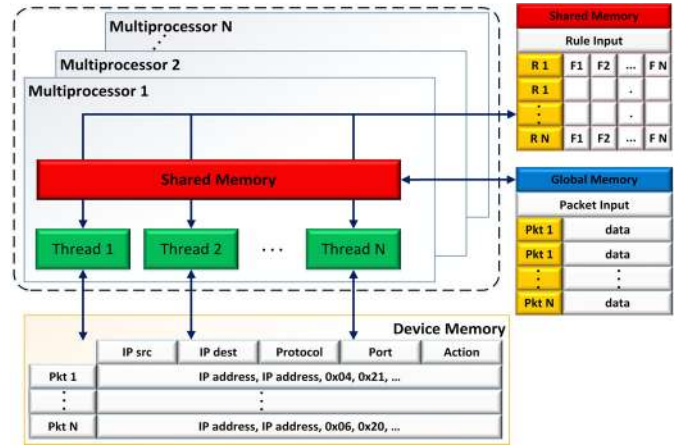


Fig. 3. The memory storage for filter-sets and packet data on CUDA device.

version of TSS and TPS are vastly parallel programs, in which fine-grained parallelism is realized by exploiting the maximum number of concurrent threads to classify packets. Note that, since the precise number of references to the memory during the runtime is not precisely determinable, the exact number of floating-point computations per packet could not be estimated. But as a lower bound, the balanced TSS/TPS algorithms should perform at least one floating-point operation for each packet classification.

Algorithm 1 represents the main steps of implementing TSS, and TPS algorithms using CUDA. Initially, the filter-set and packets are loaded from two separate files into the host memory. Two binary tries, one for storing source address prefixes of filters and the other for storing destination address prefixes of filters are constructed in host memory. Then, breadth-first traversals of these two trees are stored in host memory. Next, after allocating the required memory, these data, which reveal tree structures, filters of filter-set, and packets are transferred from the system memory to the global memory of GPU. In addition to these items, corresponding to each packet a specific space, denoted by a result, is allocated in the global memory of GPU to store its flow number.

Algorithm 1: The Parallel Form of TSS and TPS Algorithms

- 1: $Global\ Memory \leftarrow Host\ Memory (Tree, structure$
 $Packets, Filters)$
 - 2: $p \leftarrow ReadPacket(threadIdx)$
 - 3: $BMR \leftarrow Classification(p, tuples)$
 - 4: $Result[threadIdx] \leftarrow BMR$
 - 5: $HostMemory \leftarrow Result$
-

As shown in the second step of the pseudo-code of Algorithm 1, the proposed form for parallel kernel is very fine grain. That is, each thread picks a packet from the pool of packet in the global memory and classifies it according to the third step of the process. The output of the calcification is the identification number of the best-matched filter. As shown in Algorithm 1, in the fourth step of the pseudo-code, this result is stored in a pre-allocated memory space corresponding to

each classified packet. After classifying all packets, the result is transferred from device storage to host storage. Note that after transferring the result to hot memory, the algorithm frees the allocated space in device memory.

A. Scenario 1: Using Global Memory

In the proposed mechanism for parallelizing the packet classification operation first, the source tree and destination tree corresponding to the filters are constructed by the CPU. Then the remaining data structure including two trees, the filter-set, the packet header, and the result array is transferred from the system memory to the pre-specified memory module on GPU. The reason for keeping this data in global memory is that they are reachable to all threads.

It should be noted that the Result array stores the best matching filter of each packet within a corresponding index of it. These operations are illustrated in lines 1 and 2 of Algorithm 2. Each thread classifies the maximum number of packets equal to $\frac{\text{the number of headers}}{3072}$. The number of threads used is 3072. Finally, each thread stores the result of the classifying each packet in the position corresponding to the index of that packet in the output array. This operation is illustrated in lines 3 through 8 of Algorithm 2. The classification process is performed in parallel with all threads. When all threads are done, the output array is transferred to the host memory. This operation is illustrated in lines 9 and 10 of Algorithm 2.

Algorithm 2: Global Memory Scenario

Input : *rulesR, header H*
Output: *rule indexes I for each header h ∈ H*
 1 *array results[number of headers];*
 2 *Global Memory ← Host Memory(Source Tree structure, Destination Tree structure, H, R, results);*
 3 *tid ← threadIdx;*
 4 **while** *tid < number of headers do*
 5 *p ← Readheader(tid);*
 6 *BMR ← Classification(p, tree);*
 7 *Result[tid] ← BMR;*
 8 *tid ← tid + threadIdx;*
 9 *__syncthreads();*
 10 *HostMemory ← Result*
 11 **end While**

B. Using Shared Memory

In this study, three different scenarios of shared memory-based parallel packet classification are presented. These three methods have commonalities, which will be discussed below.

The access rate and the capacity of memory modules of GPU are reversely related. That is, by increasing the access-rate, the capacity of the memory module is decreased. In the classification of the packet on GPU, two essential parts of the dataset, including filters and headers, should be stored on the most suitable modules of the memory hierarchy of GPU. Each thread accesses the header of the packet once, but it

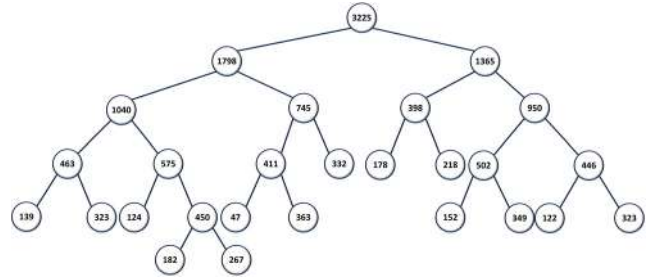


Fig. 4. Creating ACL2_1K destination tree with a 6-Block for each processing kernel.

requires access to the filters set multiple times. Therefore, keeping the filter-set in a fast memory results in a considerable decrease in the classification time of packets. On the other hand, the filter-sets in network processors of core routers include millions of entries. Regarding those two major concerns, it seems that the shared memory of GPUs with 48KB of storage is the best module for keeping the filter-set. This memory module is so advantageous since all threads in a CUDA block can simultaneously access to the content of this memory.

Each SM uses shared memory. This memory space is split between blocks in the GPU. The storage space is so small that if all this memory is allocated to a block, the source and destination trees created corresponding to the filters, will not be included in the associated memory. The solution used in this article is to break the tree from different levels. To do this, it is first estimated the amount of data that should be communicated between the CPU and the GPU. This is based on the amount of space required for tree storage and storage space. Then the blocking policy is specified. Based on the policy selected, the amount of memory allocated to each block is specified. The number of nodes in a block can be specified and used as a threshold. Then the breadth-first traversal is done on the tree and each node encountered along the path of the traversal is taken as the root of a separate tree. Then, the algorithm gets the number of allocated nodes in this new binary-tree. If the number of allocated nodes is less than the given threshold, a cut is done on the current node, and the corresponding subtree is separated from the main tree. This is done as long as the breadth-first traverse continues. It should be noted that this traverse does not include beneath cut trees. These operations are illustrated in lines 1-2 of the algorithm in Algorithm 3. Fig.4 demonstrates how to cut the ACL2_1K (Access Control Lists type 2 with 1K Rules) destination tree. The cutting process is explained below.

By estimating the number of memory transfers between the memory of the system and device, and choosing the maximum number of blocks, six blocks are defined in the processing kernel. By choosing a 6-block policy for each processor, each block reaches 8kB of shared memory, holding $512 = 8kB / 16B$ nodes. Note that the amount of space needed to store every node of trees is 16 bytes. The numbers in each node represent the sum of all nodes to the left and right of that node plus the number of filters in that node. By cutting on the target tree of ACL2_1K, the leaf nodes in Fig. 4 are each considered as the root of a separate tree. As can be seen, the number of nodes in all newly created trees is less than 512.

Algorithm 3: Shared Memory Scenarios

Input : $rulesR, headers H$
Output: $rule indexes I$ for each header $h \in H$

- 1 Source tree break into Subtrees;
- 2 Destination tree break into Subtrees;
- 3 array $O[|rules|][|headers|]$;
- 4 **for** $i \leftarrow 0$ **to** $|headers| - 1$ **do**
- 5 $O[founruleoftree][i] += 1$
- 6 **end for**
- 7 Global Memory \leftarrow Host Memory(subtrees, H, O);
- 8 **for** $i \leftarrow 1$ **to** $ceilf(|subtrees|/|blocks|)$ **do**
- 9 $tid \leftarrow threadIdx$;
- 10 $_shared_Tree$ tree[maximum size of the subtrees];
- 11 **If** (subtree(blockIdx.x*i) \neq Null) **then**
- 12 **while** $tid <$ size of subtree(blockIdx.x*i) **do**
- 13 $tree[tid] = subtree(blockIdx.x*i)$;
- 14 $tid \leftarrow tid + threadIdx$
- 15 **end While**
- 16 $_syncthread()$
- 17 **end if**
- 18 **for** $j \leftarrow 0$ **to** $ceilf(|headers|/dimBlock)$ **do**
- 19 $index \leftarrow threadIdx.x + j * dimBlock$
- 20 **If** ($index <$ $|headers|$) **then**
- 21 $O[found rule of tree][index] += 1$
- 22 **end if**
- 23 **end for**
- 24 **end for**
- 25 array results[|headers|];
- 26 Global Memory \leftarrow Host Memory($R, H, Results$);
- 27 Shared Memory \leftarrow Global Memory(R);
- 28 $tid \leftarrow threadIdx$;
- 29 **while** $tid <$ $|headers|$ **do**
- 30 $p \leftarrow Readheader(tid)$;
- 31 $BMR \leftarrow Classification(p)$;
- 32 $Result[tid] \leftarrow BMR$;
- 33 $tid \leftarrow tid + threadIdx$;
- 34 **end While**
- 35 Host Memory \leftarrow Result

With this type of cutting, the nodes on the top of the tree do not include any newly created trees. To solve this problem, a straightforward classification operation is done for each incoming packet. For this purpose, a two-dimensional array is created first. The number of packets and the size of the filter-set specifies the size of two corresponding dimensions of this array. All cells of this array are preset, first. The corresponding cells of this array are incremented in each step for the filters of the matched tuples in the traversal path. This operation is illustrated in lines 3 through 6 of Algorithm 3.

Then the new trees, the headers of the packets, along with this two-dimensional array, are transferred to the machine's global memory (line 6-7 of Algorithm 3). The next line of Algorithm 3 is executed when the number of trees created is likely to exceed the number of available blocks. In this case, the shared memory of the blocks must be emptied and reused for the remaining trees. The trees are transferred to their own

block shared memory. Each thread gets one or more nodes from the tree corresponding to its block and copies it to its block memory.

The number of nodes that each thread copies in its shared memory is obtained by $\frac{\text{the size of subtree}}{\text{dimBlock}}$ relation. Here it is necessary to make a complete copy of the tree in order to proceed to the next steps. This operation is illustrated in lines 9 through 17 of Algorithm 3.

Each thread picks $\frac{\text{the number of headers}}{\text{dimBlock}}$ of packets from the global memory and performs search operations on the tree in its block. The algorithm records the filters in the tuples encountered in the traversal path in the 2D array in the global-memory module. When all the trees are transferred to the shared-memory module and the initial search operation is performed on all packets on the trees, the two-dimensional array created in the block is transferred to the host memory. This operation is illustrated in lines 18 to 23 of the algorithm in Algorithm 3.

Then a new kernel function is created to perform the second phase of the search, which is an exhaustive search. Then the structure of the packet-headers, the filter-set, the two-dimensional array completed in the previous step, along with a new vector, and the number of test packets is copied to the pre-assigned memory of GPU to store the classification results. This operation is illustrated in lines 25 and 26 of Algorithm 3.

The linear search in the TSS is done on the filters with the corresponding value for the packet in the two-dimensional array being one or two. In the pruned tuple space algorithm, this value should be two. The reason for this difference is that the linear search in the TSS is run on the filters in the tuples that exist in the path of the source or destination tree traversal, or both, while in the pruned tuple space algorithm, the linear search on the filters which exist the ways of traversing source and destination tree. Finally, when all threads are processed, the Result array is transferred to the global storage of GPU. This transfer is performed by the functions of lines 28 to 35 of Algorithm 3. The differences between the scenarios implemented in the shared memory are presented below. Line 27 will also be executed if the linear search used in both of these is executed on the shared memory.

C. Scenario 2

This method tries to use the maximum number of blocks to decrease the number of sweeps between the CPU and GPU. Here, the key idea is to break down the tree more. This will reduce the workload of the threads in the block. Linear search is also performed on global memory in this scenario.

D. Scenario 3

The difference of this scenario with the previous one is to perform an exhaustive search on the shared memory of the block. This scenario copies the filters to the shared memory to access each thread earlier.

TABLE V
ANALYSIS PARAMETERS [1]

Parameter	Description
Q	Number of cores per core group (SMs)
L	Time for a slow global memory access
P	Total number of processors (cores)
T_1	Total number of operations in the program (work)
M	Number of global memory transactions
τ	Number of threads per core
n	Number of filters in classifier
a	Number of packets
B_a	Number of active thread blocks on each SM
B_r	Number of requested thread blocks for parallel algorithm
n_T	Number of threads on each block

V. ANALYZE USING THE CALIBRATED ASYMPTOTIC FRAMEWORK

Analytically estimating the complexity of parallel kernels is of great importance in designing resource-efficient algorithms. Recently several methods have been offered that estimate the efficiency of parallelized algorithms on many-core machines like GPUs [21]–[24].

Recently, a comprehensive method is presented for analyzing the complexity of intricate parallel algorithms on many-core computing systems like GPUs [1]. In this method, to estimate the efficiency of parallel kernels, different parameters, including the running time of the sequential algorithm, the number of SPs, the data communication time, and the number of concurrently running threads on each SP are considered. Table V, introduces the parameters that are used in this framework [1].

In our analysis model, the total time of running the parallel kernel on the graphic processor according to the proposed scenarios is obtained by equation (1):

$$Time_{total} \propto \left[\frac{a}{n_T} \right] \times \max \left(T_1, \frac{M \times L}{\tau} \right) \times \left[\frac{QB_r}{PB_a} \right] \times \frac{1}{P} \tag{1}$$

In the above equation, T_1 and $\frac{ML}{\tau}$ represent the computational burden of the sequential deployment of the algorithm and the number of simultaneous accesses to the storage, correspondingly. The former depends on the nature of the sequential algorithm and the latter depends on the kernel policy that dictates how to deposit the data structure of the algorithm on the hierarchy of memory modules of the GPU. Given that P/Q represents the number of SMs, if $B_r > B_a \times (\frac{P}{Q})$, to completely execute the kernel, it is required to re-fill the shared memory of defined blocks $\left[\frac{QB_r}{PB_a} \right]$ times. The $(M \times L)$ term in equation (1), represents the memory complexity of the kernel.

A. Parameters Required for the Proposed Parallel Model

According to equation (1), the parameters of our analysis in different scenarios have different values regarding the size of the filter-set, the number of the SMs and SPs, and the type of used memory modules. Table VI shows the value of these parameters for the proposed parallel model.

TABLE VI
VALUE OF REQUIRED PARAMETERS

Q	L	P	T_1	n	B_a	B_r	n_T	τ
192	100	384	$O(W^2) + O(N + N)$	1024	6	24	256	8

TABLE VII
THE MEMORY COMPLEXITY

Scenario	Memory Complexity
Scenario 1	$O(W^2) * L + O(N + N) * L$
Scenario 2	$O(N) * L + O(W^2) + O(N + N) * L$
Scenario 3	$O(N) * L + O(W^2) + O(N + N)$

The parameters B_r and B_a show the number of reconstructed subtrees and the number of exploited functional blocks in each SM.

The value of parameter T_1 is computed to the extent of $O(W^2) + O(N + N)$ given the searching process of the algorithm and the way of comparing the fields of each filter of filter-set with the values extracted from their corresponding fields of the packet header. Parameter τ is computed according to the following equation and reflects the computational load for each SP.

$$\tau = \frac{n_T \times \frac{P}{Q} \times B_a}{P} \tag{2}$$

B. The Complexity of the Proposed Scenarios

The complexity of transferring the data from the global memory to the shared memory is $O(n)$. Also, the total complexity of searching the tree to find all matching filters is $O(w^2) + O(n + n)$. Note that the type and combination of memory modules used for storing the essential data determine the worst-case memory complexity of each scenario. The memory complexity of all scenarios is calculated and shown in Table VII. The linear search in the second and third scenarios is performed in both global and shared memories. Hence, the memory complexity of the third scenario is the lowest as compared to others. Based on the parameters of the parallel model, Table VIII shows the complexity of the proposed kernels on a CUDA device with specifications represented in Table IX.

Fig. 5 shows the complexity diagram of the various scenarios presented for parallel packet classification using the TSS algorithm. The obtained complexity is calculated according to the parameter values of Table VII and the scenarios of Table IX. Given the complexity results, it is predicted that the third, first, and second scenarios will yield the best results, respectively.

VI. IMPLEMENTATION AND EVALUATION

The first part of this section is focused on the technical details of the implementation of the parallel kernels and the ClassBench tool [16]. Finally, the results of the implementation of both algorithms on different filter-sets are compared.

TABLE VIII
THE COMPUTATIONAL COMPLEXITY OF THE PRESENTED PARALLEL SCENARIOS ON THE FILTER-SET UNDER EVALUATION
BASED ON THE CALIBRATED ASYMPTOTIC FRAMEWORK

Scenario	Complexity
<i>Scenario 1</i>	$Time_{total} \propto \left\lceil \frac{a}{256} \right\rceil \times \max \left(O(W^2) + O(N + N), \frac{O(W^2) * L + O(N + N) * L}{8} \right) \times \left\lceil \frac{QB_r}{PB_a} \right\rceil \times \frac{1}{P}$
<i>Scenario 2</i>	$Time_{total} \propto \left\lceil \frac{a}{256} \right\rceil \times \max \left(O(W^2) + O(N + N), \frac{O(N) * L + O(W^2) + O(N + N) * L}{8} \right) \times \left\lceil \frac{QB_r}{PB_a} \right\rceil \times \frac{1}{P}$
<i>Scenario 3</i>	$Time_{total} \propto \left\lceil \frac{a}{256} \right\rceil \times \max \left(O(W^2) + O(N + N), \frac{O(N) * L + O(W^2) + O(N + N)}{8} \right) \times \left\lceil \frac{QB_r}{PB_a} \right\rceil \times \frac{1}{P}$

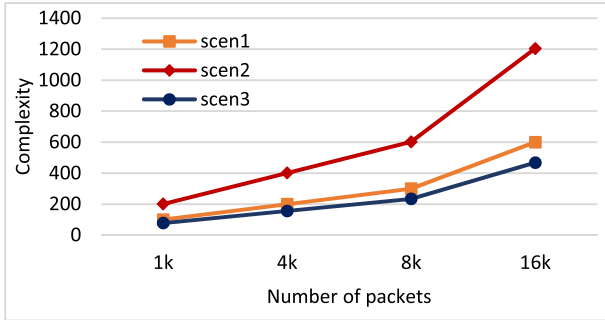


Fig. 5. The complexity of different parallel classification scenarios using TSS.

A. Implementation

The required synthetic data is produced using ClassBench. It is used to generate required filter-sets and packets based on input parameter files [25], [26]. The presence of this tool resolves the need for collecting real packet profiles, and real filter-sets of Firewalls (FW), IP Chains (IPC), and Access Control Lists (ACL). A different number of filters, related to ACL2 and FW2, with $N = 1k$ filters, was generated using ClassBench and used in all of our experiments. Corresponding to each of the filter-sets above, different profiles, including 1k to 64k synthetic packet headers were generated using ClassBench and used in all experiments. The parallelized version of TSS and TPS algorithms were developed by C programming language. The CUDA 9 platform was used to implement the GPU-based programs. The provided scenarios were tested on a CUDA device with specifications represented in Table IX.

B. Evaluation

In this section, the proposed scenarios are examined from different aspects of performance, such as classification time and throughput. The classification throughput shows the number of packets that are classified per unit time. The proposed scenarios were run the algorithms on the filter-sets ten times, and then the average of the aforementioned parameters was calculated. All times quoted are in a millisecond.

Fig. 6 shows the packet classification times of 1k to 16k of packets using the parallel TSS algorithm and the parallel TPS algorithm. Charts a and b show the results of the ACL filter-set, and the c and d charts show the result of the FW filter-set.

According to the results of the two filter-sets, the lowest classification time belongs to the third scenario. Because in

TABLE IX
SPECIFICATION OF THE SYSTEM

Device	Specification
VGA	Nvidia GeForce GT 645M(Keppler) / 2GB DDR3
CUDA cores	384
Graphics clock	708 MHz
Memory bandwidth	28.80 GB/s
Shared memory	48 KB
CPU	Intel®Core™i7-3630QM
RAM	32 GB
Operation System	Windows 7 Ultimate,64-bit (SP1)

this scenario, a binary search is performed to match all of the packet fields in shared memory. The first scenario, where the tree is wholly held in global memory, is faster than the second scenario where the tree is formed from small sub-trees in shared memory. The results obtained from the implementation confirm the predictions made in the previous section proposed by scenarios.

After traversing two binary trees according to the corresponding fields of the packet, all of the visited tuples and all of the commonly visited tuples during the traversal of the respectively TSS, and TPS algorithms are inspected to find the best-matched filter. Since the number of inspected tuples in TPS is much lower than that of TSS, the packet classification time by the former would be considerably lower than the latter. The plots of classification time of the ACL and the FW datasets in Fig. 6 shows that in all scenarios, the TPS is faster than the TSS algorithm. For example, in the second scenario, for classifying 16K of packets, the speedup of TPS and TSS to the sequential version of the algorithms is 9.28 and 6.76, respectively. Fig. 7 shows the throughput of the proposed scenarios of the parallelization of the TSS and the TPS algorithms. Bar chart (a) shows the Throughput of ACL filter-set and bar chart (b) shows that of the FW filter-set per kilo packet per second. The Throughput of the TPS algorithm is more than TSS in all scenarios and filter-sets.

The main reason for the higher throughput of the TPS algorithms as compared to the TSS algorithm is that as compared with the latter, the former inspects a lower number of tuples for finding the best-matched filter at the final step of classifying packets. Among the scenarios presented, the third, first, and second scenarios have the highest amount of throughput, respectively. The highest pass rates in the ACL and FW filter-sets are 1028.28 and 926.46 KPPS, respectively, which are achieved in the third TPS scenario. This result introduces the third scenario as the best one for the real-time classification of the vehicular network packet.

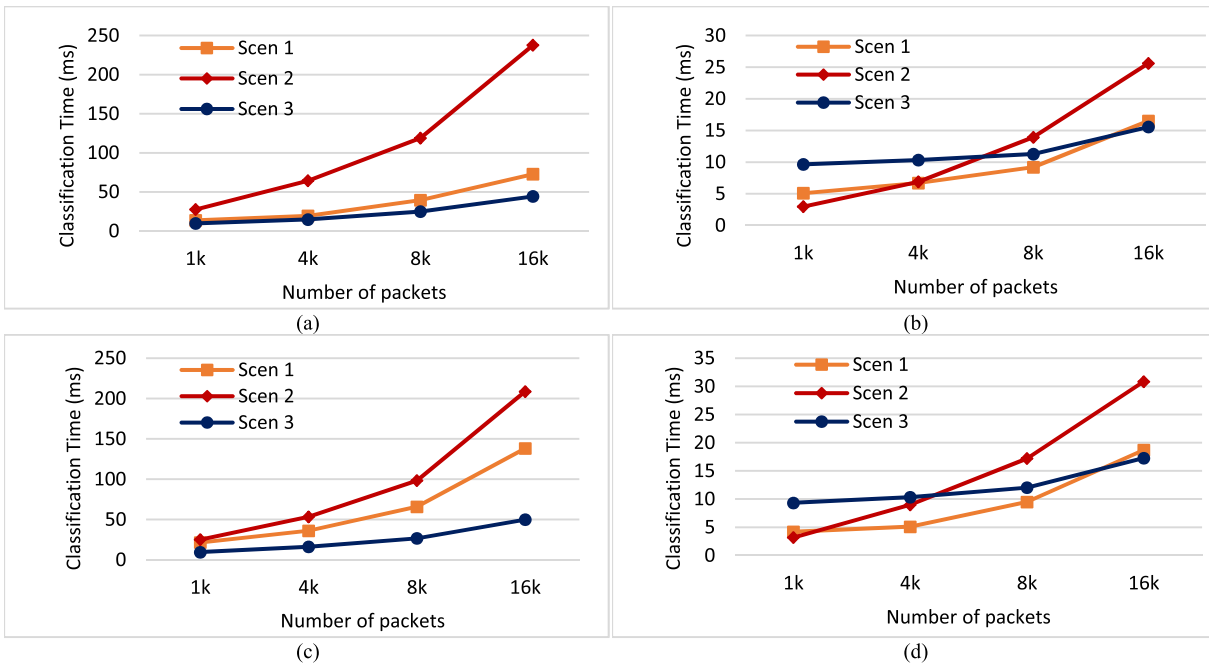


Fig. 6. Packet classification time in different scenarios and filter-sets. (a) TSS-ACL. (b) TPS-ACL. (c) TSS-FW. (d) TPS-FW.

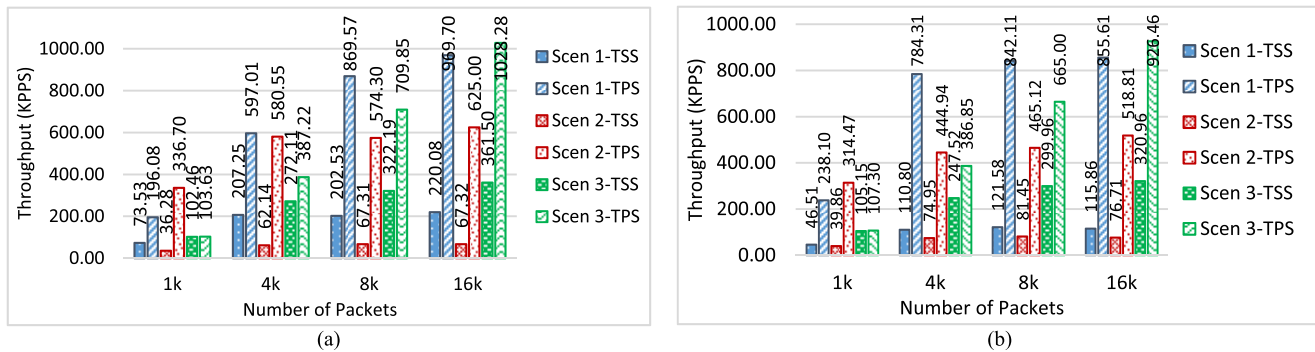


Fig. 7. Throughput of different packet classification scenarios. (a) Throughput of ACL filter-set. (b) Throughput of FW filter-set.

VII. CONCLUSION

Flow classification is considered as the key technique for accelerating the different functions of modern network systems in SDN-based IoT. The critical challenge in this regard is deploying an engine that can classify the incoming packets with speed near to the wire speed. The engine should also optimize memory usage.

Existing solutions have failed to make a fair tradeoff between the time and amount of memory consumed. This paper focuses on the TSS and TPS packet classification algorithms, which are considered suitable algorithms for parallel implementation. These algorithms work well to reduce the amount of memory consumed, but they suffer from low speed. This study has attempted to resolve this problem by parallel implementation of these algorithms. These two algorithms are implemented in the GPU in three different scenarios. To predict the classification speed of the proposed scenarios, their complexity is calculated. The computational complexity is obtained by the calibrated asymptotic model. The main parameter in evaluating the performance of parallel algorithms on the graphics processing unit is the packet classification time.

Analyzing the evaluation results show that the different modes of memory utilization in GPU have a significant effect on the speed of packet classification by TSS and TPS algorithms. Also, comparing the results of the formal complexity analysis and the corresponding experimental results confirms the efficiency of the hybrid scenario in parallelization of the TSS and TPS algorithms.

The present study would be extended by parallelizing the parallelizable algorithms on a cluster of GPUs. Without any doubt, the design of efficient GPU cluster-based parallel classifier engines requires a comprehensive analytical platform that can precisely estimate the effect of specific parameters on the complexity of kernels.

REFERENCES

- [1] M. Abbasi and M. Rafiee, "A calibrated asymptotic framework for analyzing packet classification algorithms on GPUs," *J. Supercomput.*, vol. 75, no. 10, pp. 6574–6611, Oct. 2019.
- [2] E. Benalia, S. Bitam, and A. Mellouk, "Data dissemination for Internet of vehicle based on 5G communications: A survey," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 5, p. e3881, May 2020.

- [3] S. Garg, K. Kaur, G. Kaddoum, S. H. Ahmed, and D. N. K. Jayakody, "SDN-based secure and privacy-preserving scheme for vehicular networks: A 5G perspective," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8421–8434, Sep. 2019.
- [4] E. Khaledian, N. Movahedinia, and M. Khayyambashi, "Spectral and energy efficiency in multi hop OFDMA based networks," in *Proc. 7th Int. Symp. Telecommun. (IST)*, Sep. 2014, pp. 123–128.
- [5] L. Huo, D. Jiang, and H. Qi, "A security traffic measurement approach in SDN-based Internet of Things," in *Proc. Int. Conf. Simulation Tools Techn.*, 2019, pp. 146–156.
- [6] K. Kaur, S. Garg, G. Kaddoum, E. Bou-Harb, and K.-K.-R. Choo, "A big data-enabled consolidated framework for energy efficient software defined data centers in IoT setups," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2687–2697, Apr. 2020.
- [7] J. Wee, J.-G. Choi, and W. Pak, "Wildcard fields-based partitioning for fast and scalable packet classification in vehicle-to-everything," *Sensors*, vol. 19, no. 11, p. 2563, Jun. 2019.
- [8] D. E. Taylor, "Survey and taxonomy of packet classification techniques," *ACM Comput. Surv.*, vol. 37, no. 3, pp. 238–275, Sep. 2005.
- [9] S. Zhou, S. G. Singapura, and V. K. Prasanna, "High-performance packet classification on GPU," in *Proc. IEEE High Perform. Extreme Comput. Conf. (HPEC)*, Waltham, MA, USA, 2014, pp. 1–6, doi: 10.1109/HPEC.2014.7041005.
- [10] P. Du, R. Weber, P. Luszczek, S. Tomov, G. Peterson, and J. Dongarra, "From CUDA to OpenCL: Towards a performance-portable solution for multi-platform GPU programming," *Parallel Comput.*, vol. 38, no. 8, pp. 391–407, Aug. 2012.
- [11] A. Nottingham and B. Irwin, "GPU packet classification using OpenCL: A consideration of viable classification methods," in *Proc. Annu. Res. Conf. South Afr. Inst. Comput. Scientists Inf. Technologists (SAICSIT)*, 2009, pp. 160–169.
- [12] S. Han, K. Jang, K. Park, and S. Moon, "PacketShader: A GPU-accelerated software router," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 195–206, 2010.
- [13] C.-L. Hung, H.-H. Wang, S.-W. Guo, Y.-L. Lin, and K.-C. Li, "Efficient GPGPU-based parallel packet classification," in *Proc. IEEE 10th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Nov. 2011, pp. 1367–1374.
- [14] K. Kang and Y. S. Deng, "Scalable packet classification via GPU metaprogramming," in *Proc. Design, Autom. Test Eur.*, Mar. 2011, pp. 1–4.
- [15] S. Zhou, Y. R. Qu, and V. K. Prasanna, "Multi-core implementation of decomposition-based packet classification algorithms," *J. Supercomput.*, vol. 69, pp. 34–42, Jun. 2014.
- [16] M. Varvello, R. Laufer, F. Zhang, and T. V. Lakshman, "Multi-layer packet classification with graphics processing units," presented at the 10th ACM Int. Conf. Emerg. Netw. Exp. Technol., Sydney, NSW, Australia, 2014.
- [17] J. Zheng, D. Zhang, Y. Li, and G. Li, "Accelerate packet classification using GPU: A case study on HiCuts," in *Computer Science and Its Applications*, vol. 330, J. J. Park, I. Stojmenovic, H. Y. Jeong, and G. Yi, Eds. Berlin, Germany: Springer, 2015, pp. 231–238.
- [18] V. Srinivasan, S. Suri, and G. Varghese, "Packet classification using tuple space search," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 4, pp. 135–146, Oct. 1999.
- [19] NVIDIA. (Mar. 2015). *NVIDIA CUDA Compute Unified Device Architecture Programming Guide, Version 6.5*. [Online]. Available: http://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf
- [20] A. R. Brodtkorb, T. R. Hagen, and M. L. Sætra, "Graphics processing unit (GPU) programming strategies and trends in GPU computing," *J. Parallel Distrib. Comput.*, vol. 73, no. 1, pp. 4–13, Jan. 2013.
- [21] M. Amaris, D. Cordeiro, A. Goldman, and R. Y. D. Camargo, "A simple BSP-based model to predict execution time in GPU applications," in *Proc. IEEE 22nd Int. Conf. High Perform. Comput. (HiPC)*, Dec. 2015, pp. 285–294.
- [22] Y. Deng, X. Jiao, S. Mu, K. Kang, and Y. Zhu, "NPGPU: Network processing on graphics processing units," in *Theoretical and Mathematical Foundations of Computer Science*. Berlin, Germany: 2011, pp. 313–321.
- [23] K. Nakano, "Simple memory machine models for GPUs," in *Proc. IEEE 26th Int. Parallel Distrib. Process. Symp. Workshops PhD Forum*, May 2012, pp. 794–803.
- [24] K. Nakano, "The hierarchical memory machine model for GPUs," in *Proc. IEEE Int. Symp. Parallel Distrib. Process., Workshops Phd Forum*, May 2013, pp. 591–600.
- [25] D. E. Taylor and J. S. Turner, "Classbench: A packet classification benchmark," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Societies*, Mar. 2005, pp. 2068–2079.
- [26] M. Abbasi, R. Tahouri, and M. Rafiee, "Enhancing the performance of the aggregated bit vector algorithm in network packet classification using GPU," *PeerJ Comput. Sci.*, vol. 5, p. e185, Apr. 2019.



Mahdi Abbasi is currently an Associate Professor with the Department of Computer Engineering, Bu-Ali Sina University, Hamedan, Iran. He is also the Director of the Computer Architecture Research Laboratory, Bu-Ali Sina University. His areas of research interests include computer architecture, network embedded systems, multi-objective optimization and control, distributed computing, the IoT, and signal processing.



Ali Najafi received the M.S. degree in information technology from Bu-Ali Sina University, Hamedan, Iran, in 2016. His research interests include computer networks, parallel and distributed computing, and GPU programming.



Milad Rafiee is currently a Lecturer and a Research Assistant with the Department of Computer Engineering, Bu-Ali Sina University, Hamedan, Iran. His research interests include computer networks, the IoT, software-defined networking, parallel and distributed computing, and GPU programming.



Mohammad R. Khosravi is currently with the Department of Computer Engineering, Persian Gulf University, Iran. His main interests include signal and image processing, computer networks, and distributed computing.



Varun G. Menon (Senior Member, IEEE) is currently an Associate Professor with the Department of Computer Science and Engineering, Kerala, India. His main interests include security and privacy, computer networks, and cyber-physical systems.



Ghulam Muhammad (Senior Member, IEEE) is currently a Professor with the Department of Computer Engineering, King Saud University (KSU), Riyadh, Saudi Arabia. His research interests include multimedia data processing, healthcare systems, and machine learning.



Full length article

Efficient equalisers for OFDM and DFrFT-OCDM multicarrier systems in mobile E-health video broadcasting with machine learning perspectives

Hani H. Attar^a , Ahmad A.A. Solyman^b , Abd-Elnaser Fawzy Mohamed^c , Mohammad R. Khosravi^{d e} ,
Varun G. Menon^f  , Ali Kashif Bashir^g , Pooya Tavallali^h 

^a Department of Energy Engineering, Zarqa University, Jordan

^b Department of Electrical and Electronics Engineering, Istanbul Gelisim University, Turkey

^c Communication Department, Bilbeis Higher Institute for Engineering, Egypt

^d Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

^e Department of Electrical and Electronic Engineering, Shiraz University of Technology, Shiraz, Iran

^f Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kochi, India

^g Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, United Kingdom

^h Department of Electrical Engineering and Computer Science, University of California, Merced, CA 95343, USA

Received 28 February 2020, Revised 5 July 2020, Accepted 15 July 2020, Available online 18 July 2020, Version of Record 24 July 2020.



Show less 

 Share  Cite

<https://doi.org/10.1016/j.phycom.2020.101173> 

[Get rights and content](#) 

Abstract

Recently, the orthogonal frequency-division multiplexing (OFDM) system has become an appropriate technique to be applied on the physical layer in various requests, mainly in wireless communication standards, which is the reason to use OFDM within mobile wireless medical applications. The OFDM with cyclic prefix (CP) can compensate lacks for the time-invariant multi-path channel effects using a single tap equaliser. However, for mobile wireless communication, such as the use of OFDM in ambulances, the Doppler shift is expected, which produces a doubly dispersive communication channel where a complex equaliser is needed. This paper proposes a low-complexity band LDL^H factorisation equaliser to be applied in mobile medical communication systems. Moreover, the discrete fractional Fourier transform (DFrFT) is used to improve the communication system's performance over the OFDM. The proposed low-complexity equaliser could improve the OFDM, and the DFrFT-orthogonal chirp-division multiplexing (DFrFT-OCDM) system's performance, as illustrated in the simulation results. This proves that the recommended system outperforms the standard benchmark system, which is an essential factor as it is to be applied within mobile medical systems.

Introduction

Mobile wireless communication systems for E-health applications have received more attention recently with the goal to achieve a mobile hospital and patient monitoring system. Accordingly, the mobile wireless communication system, including video broadcasting features, is urgently needed in such applications. The orthogonal frequency-division multiplexing (OFDM) is the base for several communication systems such as, European digital video broadcasting systems like Digital Video Broadcasting-Terrestrial (DVB-T), DVB-Second Generation Terrestrial (DVB-T2), DVB Handheld (DVB-H), Long-Term Evolution (LTE), and fifth generation (5G) mobile communication systems. The popularity of OFDM systems is based on its ability to compensate the effect for the time-invariant channel matrix. However, the OFDM loses its optimality against intercarrier interference (ICI) due to Doppler shift (doubly dispersive channel) or carrier frequency offset; accordingly, the system will be in need of sophisticated equalisers [1], [2]. In [3], [4], the discrete Fourier transform (DFT) was replaced by the discrete fractional Fourier transform (DFrFT) for multicarrier systems, which resulted in decreasing the Doppler frequency spread's effect, benefiting from the DFrFT subcarrier's chirped nature that mitigates the Doppler shift. As such, the ICI was reduced.

While DFrFT gives a more improved performance than DFT under the doubly dispersive fading channel, there is still a need for a complex equaliser [5]. Simple equalisers were proposed in [6], [7], [8], wherein [6] least-squares problems (LSQR) algorithm was offered to solve the matrix inversion iteratively; accordingly, no matrix inversion is needed, which simplifies the equaliser. In [7], the equaliser was simplified by using a banded matrix, while in [8], a new approach is proposed, which is based on both a banded matrix and the LDL^H factorisation algorithm.

Unlike the aforementioned simple equalisers which were applied with OFDM, this paper proposes the Orthogonal Chirp Division Multiplexing (DFrFT-OCM) systems, and then combines the simple suggested equaliser under a time-variant multi-path channel, which is deemed to be suitable for a medical mobile video broadcasting system. Furthermore, the DFrFT-OCM system was introduced in detail, including the way it is able to replace the DFT on the OFDM, and primarily, the simple equaliser has been added with DFrFT-OCM to fit the mobile medical applications. Moreover, the doubly dispersive channel details, with their effects on the OFDM and the DFrFT-OCM system's performance, will be outlined. The equalisation challenges will be specified, then an assessment between some known complex equalisers will be delivered to evaluate the equalisers' behaviour when improving the systems. The suggested simple equalisation methods based on the LDL^H factorisation algorithm is explained and presented as a practical solution for mobile medical applications.

802.11-WLAN video streaming was investigated in [9] over m-health claims, where a medical channel-adaptive fair allocation scheme was proposed to enhance the Quality of service (QoS) for IEEE 802.11 (WLAN). More recent work against real-time medical applications were explained in [10], where an adaptive video encoder compared to a real-time medical use is investigated to maximise the encoded video's quality, improve encoding rate, and to minimise the bit rate demands. In [11], an experimental set was introduced to provide mobile WiMAX video streaming performance analysis for Bandwidth on demand (BOND) services. More recent research and proposed wireless medical applications can be found in [12], [13], [14], [15]. Comprehensive knowledge regarding the structure of health monitoring and machine learning can be found in the well-cited reference book [16], where the theory and the demonstration of the health monitoring structure were presented. Recently, a lot of research has been carried out in this field, including [17], where the limitations of machine learning approaches have been investigated, and future clinical translations defined. Specific application for using machine learning within health monitoring is rapidly increasing, for example, in [18] where this technique was proposed for the early prediction of asthma attacks.

On the other hand, [19] investigates the scenario of E-health applications that apply the multi-service stream network, which concludes that the mathematical model class $G / G / 1$ – in its general case of a single-channel system – is regarded as an appropriate technique to be implemented within the E-health applications. Indeed, the short time delay and the jitter are practically suitable for E-health primarily. Moreover, the packet losses and the error rates are also considered to be suitable within E-health.

The paper is organised as follows: In Section 2, the background of the OFDM system equalisation is explained to equip the reader with a more comprehensive understanding of the research work presented. The preliminaries for this research are also stated in this section. The proposed approach is presented in Section 3. Section 4 details the

performance analysis of the proposed method, technical discussion, and deep computing/machine learning perspectives. Finally, conclusions are presented in the last section.

Section snippets

Background and preliminaries

The OFDM allows high data rates to be reliably and efficiently transmitted over the delay-dispersive channels. By dividing the transmitted signal into several narrow bandwidth sub-carriers, OFDM can mitigate the undesirable multi-path effects, mainly, the inter symbol interference (ISI) quantity in long symbol time systems. Moreover, at the beginning of each symbol, a guard period is added – termed cyclic prefix (CP) – to eliminate the expected effects of ISI over the multi-path signals' delay. ...

Proposed method

Fig.9 shows the OFDM system data flow, $\mathbf{x}_n = [\mathbf{x}_0, \mathbf{x}_1 \dots \mathbf{x}_{N_a-1}]^T$ is the data vector transmitted in the n th OFDM symbol, whilst its samples are permuted by the binary matrix $\mathbf{P} \in \mathbb{Z}^{N \times N_a}$ in the frequency domain, which allocates a data vector $\mathbf{x}_n \in \mathbb{C}^{N_a}$ to N subcarriers, with only N_a active: $\mathbf{P} = [\mathbf{0}_{N_a \times (N-N_a)/2} \mathbf{I}_{N_a} \mathbf{0}_{N_a \times (N-N_a)/2}] \mathbf{I}_{N_a}$ is an identity matrix with $N_a \times N_a$ dimensions. The vector $\mathbf{s}_n = [s_0 s_1 \dots s_N]^T$ is calculated from: $\mathbf{s}_n = \mathbf{F}^H \mathbf{P} \mathbf{x}_n$ where \mathbf{F}^H is the N -point IDFT matrix.

The time and frequency fading channel can be...

Results and discussion

In the following channel environments, the performance of uncoded bit error rate (BER) for the conventional OFDM and DFrFT-OCDM systems is studied:

- 1- Time-invariant channel....
- 2- Time-variant channel....

The QPSK modulated OFDM system under investigation has the following parameters:

$L = 8, N = 128,$ and $N_a = 96$. The communication channel simulated in this proposed work is the Rayleigh fading channel that has an exponential power delay profile and a root-mean-square delay spread of 3. The adopted carrier frequency is ...

Conclusions

In this paper, a mobile medical video streaming broadcasting system was proposed. The time and frequency fading channel with its effects on OFDM system performance were investigated. The DFrFT-OCDM MCM system was studied as an alternative MCM system that can enhance the overall MCM system performance. It was demonstrated that using simple equalisers with MCM systems was in high demand within the medical video broadcasting system because of its large symbol length, despite its simplicity. The...

CRedit authorship contribution statement

Hani H. Attar: Conceptualization, Methodology, Writing - original draft. **Ahmad A.A. Solyman:** Conceptualization, Methodology, Writing - original draft, Investigation, Supervision. **Abd-Elnaser Fawzy Mohamed:** Investigation, Visualization. **Mohammad R. Khosravi:** Resources, Supervision, Writing - review & editing. **Varun G. Menon:** Software, Resources, Writing - review & editing. **Ali Kashif Bashir:** Software, Resources, Writing - review & editing. **Pooya Tavallali:** Resources, Writing - review & editing....

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper...

Hani H. Attar received his Ph.D. from the Department of Electrical and Electronic Engineering, University of Strathclyde, United Kingdom in 2011. Since 2011, he has been working as a researcher of electrical engineering and energy systems. Dr Attar is now a university lecturer at Zarqa University, Jordan. His research interests include Network Coding, Wireless Sensor Networks, and Wireless Communications....

[Special issue articles](#) [Recommended articles](#)

References (47)

TiejunW.

Performance degradation of OFDM systems due to doppler spreading

IEEE Trans. Wirel. Commun. (2006)

Q. Huang, et al. A novel OFDM equalizer for large doppler shift channel through deep learning, in: 2019 IEEE 90th...

MartoneM.

A multicarrier system based on the fractional fourier transform for time-frequency-selective channels

IEEE Trans. Commun. (2001)

NassiriM. *et al.*

Comparative performance assessment between FFT-based and FRFT-based MIMO-OFDM systems in underwater acoustic communications

IET Commun. (2018)

Yang-SeokC.

On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels

IEEE Trans. Commun. (2001)

G. Taubock, et al. LSQR-based ICI equalization for multicarrier communications in strongly dispersive and highly mobile...

L. Guanghui, et al. Simple equalization of OFDM signal over doubly selective channels, in: Intelligent Signal...

Luca RuginiP.B. *et al.*

Low-complexity banded equalizers for OFDM systems in doppler spread channels

EURASIP J. Appl. Signal Process. (2006)

Y.E. Tan, et al. Fragility Issues of Medical Video Streaming over 802.11e-WLAN m-health Environments, in: 2006...

AntoniouZ.C.

Real-time adaptation to time-varying constraints for medical video communications

IEEE J. Biomed. Health Inf. (2018)



View more references

Cited by (18)

[EMMM: Energy-efficient mobility management model for context-aware transactions over mobile communication](#)

2021, Sustainable Computing: Informatics and Systems

Citation Excerpt :

...The geographical region is partitioned in small service area to efficient use of frequency, which is known as a cell. Fixed host (FH), base station (BS), mobile support station (MSS) and mobile host (MH) are the pillar of mobile communication [1,2]. FH used as permanent shareable data repositories and linked via high speed wired network infrastructure...

[Show abstract](#) ✓

[Quantum-enhanced OFDM signal processing: advancing 5G mobile network and IoT communication for healthcare application](#) ↗

2024, Optical and Quantum Electronics

[ML-driven optimisation of physical layer characteristics in an interweaving of ICT and metaverse](#) ↗

2023, Mobilkommunikation: Technologien und Anwendungen - 27. VDE ITG-Fachtagung

[OPTIMIZED MIMO BASED ENHANCED OFDM FOR MULTI CARRIER SYSTEM WITH 5G WAVEFORMS](#) ↗

2023, Economic Computation and Economic Cybernetics Studies and Research

[Single-Frequency Network Terrestrial Broadcasting with 5GNR Numerology Using Recurrent Neural Network](#) ↗

2022, Electronics (Switzerland)

[Mobile multimedia computing in cyber-physical surveillance services through UAV-borne Video-SAR: A taxonomy of intelligent data processing for IoMT-enabled radar sensor networks](#) ↗

2022, Tsinghua Science and Technology

[View all citing articles on Scopus](#) ↗

Hani H. Attar received his Ph.D. from the Department of Electrical and Electronic Engineering, University of Strathclyde, United Kingdom in 2011. Since 2011, he has been working as a researcher of electrical engineering and energy systems. Dr Attar is now a university lecturer at Zarqa University, Jordan. His research interests include Network Coding, Wireless Sensor Networks, and Wireless Communications.



Ahmad A. A. Solyman graduated from the University of Strathclyde, United Kingdom in 2013. His Ph.D. researches lie in Multimedia Services over Wireless Networks Using OFDM. He is currently a lecturer at the Department of Electrical and Electronics Engineering, Istanbul Gelisim University, Turkey. His research interests contain Wireless Communication Networks, and MIMO Communication Systems.



Abd-Elnaser Fawzy Mohamed is currently a researcher with Bilbeis Higher Institute for Engineering, Egypt. He received his Ph.D. in the area of Electrical and Computer Engineering from Cairo University, Egypt.



Mohammad R. Khosravi is with the Department of Electrical and Electronic Engineering, Shiraz University of Technology, Iran, and Department of Computer Engineering, Persian Gulf University, Iran. His main interests include statistical signal and image Processing, medical bioinformatics, radar imaging and satellite remote sensing, computer communications, industrial wireless sensor networks, underwater acoustic communications, information science and scientometrics.



Varun G. Menon is currently Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include sensors, IoT, fog computing and underwater acoustic sensor networks. He has completed his Ph. D in Computer Science and Engineering from Sathyabama University, India in 2017. He is also a Distinguished Speaker of ACM.



Ali K. Bashir is a Senior Lecturer at the Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. His past assignments include Associate Professor of Information and Communication Technologies, Faculty of Science and Technology, University of the Faroe Islands, Denmark; Osaka University, Japan; Nara National College of Technology, Japan; the National Fusion Research Institute, South Korea; Southern Power Company Ltd., South Korea, and the Seoul Metropolitan Government, South Korea. He received his Ph.D. in computer science and engineering from Korea University, South Korea. MS from Ajou University, South Korea and BS from University of Management and Technology, Pakistan. He is supervising/co-supervising several graduate (MS and Ph.D.) students. His research interests include internet of things, wireless networks, distributed systems, network/cyber-security, network function virtualization, etc. He has authored over 80 peer-reviewed articles. He has served as a chair (programme, publicity, and track) on top conferences and workshops. He has delivered over 20 invited and keynote talks in seven countries. He is a Distinguished Speaker, ACM; Senior Member of IEEE; Member, ACM; Member, IEEE Young Professionals; Member, International Association of Educators and Researchers, UK. He is serving as the Editor-in-chief of the

IEEE FUTURE DIRECTIONS NEWSLETTER. He is advising several start-ups in the field of STEM-based education, robotics, internet of things, and Blockchain.



Pooya Tavallali received the B.Sc. and the M.Sc. degrees in Electrical Engineering (Communication Systems) from the Department of Electrical and Electronic Engineering, Shiraz University, Shiraz, Iran, in 2013 and 2016, respectively. Since 2016, he has been a Ph.D. scholar at the Department of Electrical Engineering and Computer Science, University of California, Merced, USA. His scientific interest consists of machine learning, statistical signal and image processing, neural networks, statistical pattern recognition and optimisation algorithms.

[View full text](#)

© 2020 Elsevier B.V. All rights reserved.



All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.



[Home](#) > [Journal of Real-Time Image Processing](#) > [Article](#)

Special Issue Paper | [Published: 24 June 2020](#)

Dual-mode power reduction technique for real-time image and video processing board

[Sunil Jacob](#), [Varun G. Menon](#) , [Saira Joseph](#) & [Paramjit Sehdev](#)

[Journal of Real-Time Image Processing](#) **17**, 1991–2004 (2020)

183 Accesses | **2** Citations | [Metrics](#)

Abstract

In real-time image and video processing boards, power, speed, and area are the most often used measures for determining the performance of motion imagery applications. Due to technological advancement, power consumption has gained major attention in real-time image processing ability compared to speed. The increase in on-chip temperature due to larger power consumption has resulted in reduced operating life of chip and battery-driven devices. In this work, a new logic family has been introduced i.e., dual-mode logic (DML), which provides flexibility between the optimization of energy and delay (E-D

optimization). This gate can be switched between two modes of operation that is a static mode (CMOS-like mode), which provides low power consumption and dynamic mode, which provides high speed. Recently, power leakage has become a dominant problem due to continuous data transfer among a large number of connected devices. Thus, to reduce power leakage, a self-controllable voltage level (SVL) power reduction technique is used along with DML logic. In the SVL technique, a maximum dc voltage is provided to the active load circuit on-demand or decrease the dc supplied to the load circuit in the standby mode. Integrating DML with the SVL technique reduces power consumption as well as leakage power. A 4-bit RCA, 8-bit RCA, and 16-bit RCA are used for verifying the proposed method and comparison of performance parameters is done with a conventional circuit. Complete circuit implementation and simulation are carried out in TANNER EDA version 13 tools with operating voltage of 1 V. The proposed system is further applied to real-time image, and we obtain the finest resolution level with minimum power consumption.

This is a preview of subscription content, [access via your institution](#).

Access options

Buy article PDF

39,95 €

Price includes VAT (India)

Instant access to the full article PDF.

[Rent this article via DeepDyve.](#)

[Learn more about Institutional subscriptions](#)

References

1. Ahmad, A., Anisetti, M., Damiani, E., Jeon, G.: Special issue on real-time image and video processing in mobile embedded systems. *J. Real-Time Image Proc.* **16**(1), 1–4 (2018).
<https://doi.org/10.1007/s11554-018-0842-4>.
2. Monteiro, H. A. et al. (2017). Energy Consumption Measurement of a FPGA Full-HD Video Processing Platform. In Proceedings of 7th Workshop on Circuits and Systems Design (WCAS'17), Fortaleza, Brazil.
3. Wang, D., Zhu, D., & Liu, R. (2019). Video SAR High-speed Processing Technology Based on FPGA. In 2019 IEEE MTT-S International

Microwave Biomedical Conference (IMBioC),
Vol. 1, pp. 1–4. IEEE.

4. Wu, J. (2004). CMOS transistor design challenges for mobile and digital consumer applications. In Proceedings. 7th International Conference on Solid-State and Integrated Circuits Technology, Vol. 1, pp. 90–95. IEEE.

5. Blanco-Filgueira, B., García-Lesta, D., Fernández-Sanjurjo, M., Brea, V. M., & López, P. (2019). Deep learning-based multiple object visual tracking on embedded system for iot and mobile edge computing applications. IEEE Internet of Things Journal.

6. Guduri, M., Dokania, V., Verma, R., Islam, A.: Minimum energy solution for ultra-low power applications. *Microsyst. Technol.* **25**(5), 1823–1831 (2019)

7. Markovic, D., Wang, C.C., Alarcon, L.P., Liu, T.T., Rabaey, J.M.: Ultralow-power design in near-threshold region. *Proc. IEEE* **98**(2), 237–252 (2010)

8. Hussain, I., Singh, A., & Chaudhury, S. (2018). A Review on the Effects of Technology on CMOS and CPL Logic Style on Performance,

- Speed and Power Dissipation. In 2018 IEEE Electron Devices Kolkata Conference (EDKCON), pp. 332–336, IEEE.
-
9. Baker, R. J. (2019). CMOS: circuit design, layout, and simulation. Wiley-IEEE press.
-
10. Garg, S., Gupta, T.K.: Low power domino logic circuits in deep-submicron technology using CMOS. Engineering Science and Technology, an International Journal **21**(4), 625–638 (2018)
-
11. Kaizerman, A., Fisher, S., & Fish, A. (2012). Subthreshold dual mode logic. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, **21**(5), 979–983.
-
12. Levi, I., Fish, A.: Dual mode logic—Design for energy efficiency and high performance. IEEE access **1**, 258–265 (2013)
-
13. Jain, D. K., Jacob, S., Alzubi, J., & Menon, V. (2019). An efficient and adaptable multimedia system for converting PAL to VGA in real-time video processing. Journal of Real-Time Image Processing, 1–13.
-

14. Alzubi, J., Jacob, S., Menon, V. G., Joseph, S., & Vinoj, P. G. (2018). A top-up design for pal to vga conversion in real time video processing system. In 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT) (pp. 1–5). IEEE.

15. Garcia, P., Bhowmik, D., Stewart, R., Michaelson, G., Wallace, A.: Optimized Memory Allocation and Power Minimization for FPGA-Based Image Processing. *Journal of Imaging*. **5**(1), 7 (2019)

16. Kechiche, L., Touil, L., Ouni, B.: Toward the Implementation of an ASIC-Like System on FPGA for Real-Time Video Processing with Power Reduction. *International Journal of Reconfigurable Computing*. **2018**, 1–11 (2018)

17. Mutoh, S.I., Douseki, T., Matsuya, Y., Aoki, T., Shigematsu, S., Yamada, J.: 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS. *IEEE J. Solid-State Circuits* **30**(8), 847–854 (1995)

18. Kuroda, T., Fujita, T., Mita, S., Nagamatsu, T., Yoshioka, S., Suzuki, K., & Kinugawa, M. (1996). A 0.9-V, 150-MHz, 10-mW, 4 mm/sup 2/, 2-D discrete cosine transform core processor with variable threshold-voltage (VT)

scheme. *IEEE Journal of Solid-State Circuits*, 31(11), 1770–1779.

19. Balamurugan, V. (2015, March). Performance analysis of asynchronous dual mode logic using leakage power reduction techniques. In 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS) (pp. 1–5). IEEE.

 20. Enomoto, T., Oka, Y., Shikano, H.: A self-controllable voltage level (SVL) circuit and its low-power high-speed CMOS circuit applications. *IEEE J. Solid-State Circuits* 38(7), 1220–1226 (2003)

 21. Akashe, S., Sharma, R., Tiwari, N., & Pandey, R. (2012). Modelling of high speed low power decoder in nanometer era. In 2012 World Congress on Information and Communication Technologies (pp. 13–17). IEEE.

 22. Drazdziulis, M., & Larsson-Edefors, P. (2003). A gate leakage reduction strategy for future CMOS circuits. In ESSCIRC 2004–29th European Solid-State Circuits Conference (pp. 317–320). IEEE.
-

23. Müller, H., Unay, D.: Retrieval from and understanding of large-scale multi-modal medical datasets: A review. *IEEE Transactions on Multimedia* **19**(9), 2093–2104 (2017)

Acknowledgement

Authors would like to thank Dr. Mohammad Khosravi, Department of Computer Engineering, Persian Gulf University, Bushehr, Iran for his valuable inputs that has substantially helped in revising and improving the research paper.

Author information

Authors and Affiliations

Center for Robotics, SCMS School of Engineering and Technology, Cochin, 683576, India

Sunil Jacob

Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Cochin, 683576, India

Varun G. Menon

Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Cochin, 683576, India

Saira Joseph

Department of Mathematics and Computer Science, Coppin State University, Baltimore,

MD, 21216, USA

Paramjit Sehdev

Corresponding author

Correspondence to [Varun G. Menon](#).

Additional information

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rights and permissions

[Reprints and Permissions](#)

About this article

Cite this article

Jacob, S., Menon, V.G., Joseph, S. *et al.* Dual-mode power reduction technique for real-time image and video processing board. *J Real-Time Image Proc* **17**, 1991–2004 (2020). <https://doi.org/10.1007/s11554-020-00992-x>

Received	Accepted	Published
10 October 2019	10 June 2020	24 June 2020

Issue Date

December 2020

DOI

<https://doi.org/10.1007/s11554-020-00992-x>

Keywords

Dual-mode logic (DML)

Self-controllable voltage level (SVL)

Full adder Ripple carry adder (RCA)

Power leakage

Real-time image and video processing boards

Depth Information Enhancement Using Block Matching and Image Pyramiding Stereo Vision Enabled RGB-D Sensor

Sunil Jacob¹, Member, IEEE, Varun G. Menon¹, Senior Member, IEEE, and Saira Joseph, Member, IEEE

Abstract—Depth sensing devices enabled with an RGB camera, can be used to augment conventional images with depth information on a per-pixel basis. Currently available RGB-D sensors include the Asus Xtion Pro, Microsoft Kinect and Intel RealSense™. However, these sensors have certain limitations. Objects that are shiny, transparent or have an absorbing matte surface, create problems due to reflection. Also, there can be an interference in the IR pattern due to the use of multiple RGB-D cameras and the depth information is correctly interpreted only for short distances between the camera and the object. The proposed system, block matching stereo vision (BMSV) uses an RGB-D camera with rectified/non-rectified block matching and image pyramiding along with dynamic programming for human tracking and capture of accurate depth information from shiny/transparent objects. Here, the IR emitter generates a known IR pattern and the depth information is recovered by comparing the multiple views of the focused object. The depth map of the BMSV RGB-D camera and the resultant disparity map are fused. This fills any void regions that may have emerged due to interference or because of the reflective transparent surfaces and an enhanced dense stereo image is obtained. The proposed method is applied to a 3D realistic head model, a functional magnetic resonance image (fMRI) and the results are presented. Results showed an improvement in speed and accuracy of RGB-D sensors which in turn provided accurate depth information density irrespective of the object surface.

Index Terms—Block matching, depth sensing, disparity estimation, image pyramiding, RGB-D sensor, stereo vision.



I. INTRODUCTION

DEPTH sensing is a challenging task and is the core behind numerous indoor and outdoor applications. Various sensing devices have been employed recently for capturing accurate depth information. Precision of captured information is vital for many biomedical applications and helps in the detection and in-depth analysis of diseases like tumors [1]. RGB-D cameras are one of the few preferred modern

Manuscript received December 7, 2019; revised January 22, 2020; accepted January 22, 2020. Date of publication January 24, 2020; date of current version April 16, 2020. This work was supported by IEEE EPICS under Grant 2016-12. The associate editor coordinating the review of this article and approving it for publication was Dr. Amitava Chatterjee. (Corresponding author: Varun G. Menon.)

Sunil Jacob is with the Center for Robotics, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: suniljacob@scmsgroup.org).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: varunmenon@ieee.org).

Saira Joseph is with the Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Karukutty 683576, India (e-mail: saira_joseph@scmsgroup.org).

Digital Object Identifier 10.1109/JSEN.2020.2969324

sensing systems that capture RGB images along with per-pixel depth information [2]–[4]. Currently available RGB-D sensors capture depth information at reasonable accuracy with good data transfer rates [5]–[6]. The depth range, field of view (FoV) of depth, depth resolution and frame rate of existing RGB-D sensors are presented in table IV. The depth range accuracy depends on the distance of an object in the frame from the sensor. The accuracy of the depth information is the difference between the measured depth data and the actual depth of the reference object. The precision depends on the continuous measurement of the subsequent depth information in static condition. The accuracy and precision of the depth frame of an object also depends on the distance, colour and temperature. RGB-D cameras employ a mechanism in which they project the IR radiation in an irregular pattern of dots using an IR projector and then utilize the IR cameras to detect the infrared light bounced off the object of interest. The required depth information is captured using the offset observed on comparing the projected pattern and the recorded image. Although depth information is captured with relatively good accuracy, these devices suffer from few major limitations.

One of the major constraints with RGB-D devices is in the accurate capturing of depth information with shiny or transparent objects or objects that have an absorbing matte surface [7]– [8]. Also, there can be an interference in the IR pattern due to the use of multiple RGB-D cameras and the depth information is correctly interpreted only for short distances between the camera and the object. When multiple RGB-D sensors are used with overlapping views, IR patterns of different RGB-D cameras overlap, generating IR interference. Here, each RGB-D sensor will sense and align the pattern of other RGB-D cameras with its own IR pattern and will not be able to differentiate its own IR pattern from the other. Further, the depth quality degrades creating invalid depth pixel or black pixel with no depth information. The depth map of the image can be reconstructed comparing the RGB and the IR image. Recently, few deep learning-based methods have also been proposed with RGB-D images for human motion detection and tracking [9]. Some works have also focused on improving the clarity of image and video conversions [10]. Most of the proposed methods focus on reducing the cost of the deployed system and also on enhancement of captured images. Very few methods have focused on accurate capture of depth information especially in biomedical images [3], [11]– [13].

This paper presents a novel system, block matching stereo vision (BSMV) RGB-D that utilizes the advantages offered by block matching [14] and image pyramiding along with dynamic programming [15] for accurate capture of depth information from shiny/transparent objects. Here, an IR emitter generates a known IR pattern and the depth information is recovered by comparing multiple views of the focused object. The depth map of the block matching stereo vision RGB-D camera and the resultant disparity map are fused. This fills any void regions that may have emerged due to interference or because of reflective transparent surfaces and an enhanced dense stereo image is obtained. The proposed method is applied on a 3D realistic head model and the results are presented.

The paper is organized into three sections. Related works are presented in the next section. The proposed system along with the mathematical analysis is presented in section 3. The proposed system integrated with disparity estimation, block matching, image pyramiding and dynamic programming is explained in this section. The next section presents the results obtained from the application of the proposed method on a 3D realistic head model [16]–[19]. Conclusion and future work are presented in the final section.

II. RELATED WORKS

Few techniques have been proposed in recent years for depth information enhancement of images captured by various camera sensors. In [11], authors present an interesting technique that estimates the depth of positions that are too far away and irrelevant to any objects or plans. This is enabled with the help of an RGB-D camera and a gyroscope. An advanced and flexible color-guided autoregressive model for depth recovery from low quality images captured by

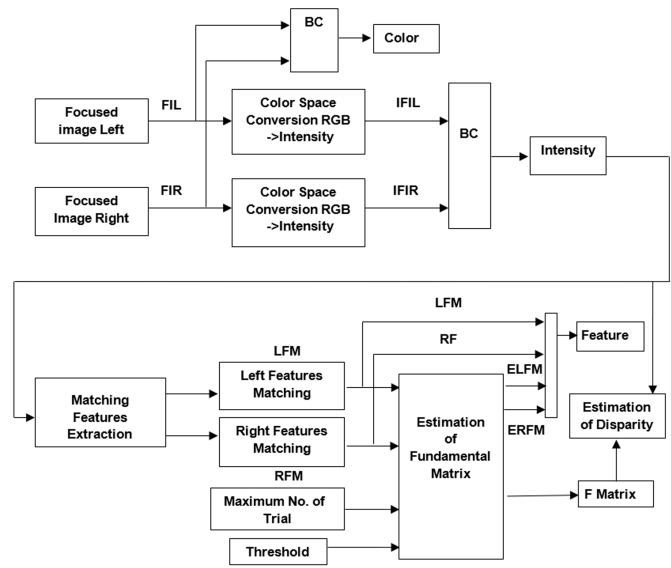


Fig. 1. System architecture.

cameras is proposed in [3]. The authors analyze the stability of the method using a linear system view. Further a parameter adapting scheme is discussed by the author for recovery of depth information. Recently, the application of RGB-D cameras has become prominent in biomedical field. Authors discuss the application of RGB-D cameras in the study of soft tissue deformation due to the bicep’s muscular activity in [12]. In [13], a depth propagation method using bilateral filtering and motion estimation is proposed. Most of the methods have certain limitations, especially in accurate detection of shiny and transparent objects. Also, some of them suffer from interference in the IR pattern due to the use of multiple RGB-D cameras and also accuracy of depth information restoration is limited. The proposed work overcomes these limitations using an RGB-D camera with rectified/non rectified block matching and image pyramiding for high speed stereo vision.

III. PROPOSED SYSTEM

Initially, the working of the system is explained with the help of an architecture diagram. Then, the mathematical modelling and analysis of the system is presented.

A. System Architecture

The architecture of the proposed system is presented in figure 1. Initially, the object of interest is captured from two different viewpoints using two RGB cameras. The focused color image pair consists of a left and right image. The focused image left (FIL) and focused image right (FIR) are connected via a bus connector (BC) and the color is extracted. Also, the color space conversion is carried out for FIL and FIR and the intensity value is extracted. Then, the matching features in each focused image is extracted using the corner detection blocks. The sum of the square difference algorithm is used to identify pixels that are common to both the focused images. The corresponding epipolar line is computed in the focused right image compared to each position in the focused left

image. On the epipolar line of the right image, the pixel that matches the pixel on the left image is identified. A quadratic polynomial is drawn such that it fits to the position of the accurately matched pixel. The polynomial curve minimum value is the final pixel location.

Here, all the matching points in left feature matching (LFM) and right feature matching (RFM) are used to compute the fundamental matrix. The maximum number of trials and the threshold value is provided for the estimation of the fundamental matrix. Eight pairs of points selected randomly from LFM and RFM along with the normalized eight-point algorithm is used to obtain the fundamental matrix. The fitness of the fundamental matrix for all points in LFM and RFM is then calculated. This value then replaces the initial fundamental matrix value. The fundamental matrix along with the calculated intensity is then used to calculate the disparity values. All the values of disparity are collected and assembled to form the disparity map.

B. Disparity Estimation

Stereo vision works on the principle of retrieving the 3D image of a structure using information of images captured from two different viewpoints. A computer compares the images while shifting the two images over each other to find the parts that match. For this, color images are first converted to gray scale using the built in `rgb2gray` function with MATLAB R2018a [20]. The shift in the pixel intensity is called the disparity and the computation output of this process is called a disparity map. The epipolar geometry of the stereo vision image is used to retrieve the relative depth information in disparity map between the two images. This gives the range between the image and the camera. However, for non-rectified images, the analysis along epipolar lines is not always aligned on the horizontal axis. The stereo images are rectified through linear transformation so that corresponding points of the two images will have same row coordinates. Rectification is a standard step used with the proposed algorithm. This step ensures that the search for matching blocks needs to be done only horizontally and not vertically during block matching.

C. Working of BSMV-RGBD

The overall working of the proposed BSMV-RGBD is discussed in three steps,

1) *Read Focused Image Left/Right Pair*: First, the object of interest is captured from two different viewpoints using two RGB cameras. The focused color image pair consists of a left and right image. From these images, color and intensity are extracted to form a grayscale image so that we have only one value (0 - 255) for each pixel. This image is used to analyze disparity map.

2) *4 × 4 Feature Block Matching*: Block matching selects a region in one image and tries to find the most similar region in the other image, using sum of absolute difference as the metric. A lower value corresponds to more similarity between the two regions. Two parameters are needed to implement block matching algorithm: size of the block and maximum disparity. In the focused left image, a 4 × 4 sub pixel feature block is

TABLE I
FREQUENTLY USED NOTATIONS

Notation	Definition
N_D	array of optimal noise disparity
I	intensity
R	optimal noise solution for $[I, d]$.
I	intrinsic matrix
$[I, d]$	boolean array threshold
D^{S-B}	baseline distance between the two RGB-D cameras from S to B
f_L	focal length_left
f_R	focal length_right
C_L	camera center_left
C_R	camera center_right
S_L	skew left
L_c	left camera
R_c	right camera
Z_d	Z depth
C_1, C_2	origin of the RGB-D camera
P_1, P_2	projection of the cross-point P on the two images

TABLE II
LIST OF ABBREVIATIONS

Abbreviation	Description
BSMV	Block Matching Stereo Vision
fMRI	Functional Magnetic Resonance Image
FoV	Field of View
IR	Infrared
RoI	Region of Interest
BC	Bus Connector
FIL	Focused Image Left
FIR	Focused Image Right
LFM	Left Features Matching
RFM	Right Features Matching
IFIL	Intensity of Focused Image Left
IFIR	Intensity of Focused Image Right
ELFM	Estimated Left Features Matching
ERFM	Estimated Right Features Matching
DP	Dynamic Programming
RHM	Realistic Head Model
NFT	Neuroelectromagnetic Forward Head Modelling
	Toolbox
BEM	Boundary Element Method

formed around a pixel. Then a search is done in the right-hand direction starting at the same coordinates in the right image. The search is carried out for the best match of the 4 × 4-pixel block. The bigger the block size, the less noisy the disparity map will be and results will be more accurate. However, it is computationally more expensive. In the proposed method, search is carried out for ±7 pixels surrounding the pixel of the focused left image. Column-wise searching is done for the rectified images.

3) *Dynamic Programming*: In the block matching process, the selected pixel has noise disparity with its own location and thus produces an image with noisy disparity. Smoothing is done to reduce noise disparity. By increasing the noise disparity sharing with surrounding sub pixels, the selected pixel shares the disparity with its surrounding sub pixel. By using four intra 16 × 16 modes we can predict one whole 16 × 16 macro block. The 16 × 16 macro block has 4 × 4 sub-blocks. These 4 × 4 sub-blocks can be predicted independently using nine intra 4 × 4 blocks. The individual prediction at intra 4 × 4 block is advantageous when the correlation at the block level is localized. 16 pixels are available in 4 × 4 sub

pixel block. This 4×4 sub pixel block helps to form the 4×4 predictive block. The 16 pixels in the 4×4 predictive block is generated using prediction mode in all the different directions using the information of all adjacent pixels. The optimal estimation of noise disparity along the noise disparity path from one block of image to another block of image can be efficiently solved using dynamic programming (DP).

The result of dynamic programming is applied individually to each block such that the optimal noise disparity structure is decomposed into suboptimal noise disparity problems. Then the optimal noise disparity value is to be expressed in terms of suboptimal noise disparity problem. The value of optimal noise disparity is then computed using a matrix structure with a bottom to up approach. This computed information is used to construct optimal noise disparity map. Next subsection presents the theoretical analysis of dynamic programming in the proposed BSMV RGB-D method.

4) *Mathematical Analysis of Dynamic Programming in BSMV-RGBD*: Initially, the optimal noise disparity structure has to be decomposed to suboptimal noise disparity problems. Here, the array of optimal noise disparity is $N_D[0 \dots n, 0 \dots n_1]$. For $1 \leq d \leq n$ and $0 \leq d \leq n_1$, the value of entry $N_D[l, d]$ will have the suboptimal noise subset with maximum combined values. The suboptimal noise subset $\{1, 2 \dots l\}$ consists of the weighting factors for intensity 'l'. Hence, the array of optimal noise disparity is given by,

$$N_D[l, d] = \max \left\{ \sum_{p \in R} n_p : R \subseteq \{1, 2 \dots d\} \sum_{p \in R} d_p \leq d \right\} \quad (1)$$

The computation of all the entries in the array $N_D[l, d]$ will lead to the $N_D[n, n_1]$ array as the suboptimal noise solution. For $[l, d]$, 'R' is a solution. If $R \subseteq \{1, 2 \dots l\}$ and $\sum_{p \in R} d_p \leq d$, then 'R' is an optimal noise solution for $[l, d]$. If 'R' is an optimal noise solution, then $\sum_{p \in R} n_p = N_D[l, d]$. Now the optimal noise disparity value in terms of suboptimal noise disparity problem is to be obtained. The initial setting for this scenario is $N_D[0, d] = 0$ for $0 \leq d \leq n_1$ and $N_D[l, d] = -\infty$ for $d < 0$. The recursive step is given by $N_D[l, d] = \max(N_D[l-1, d], n_l + N_D[l-1, d-d_l])$ for $1 \leq l \leq n, 0 \leq d \leq n_1$. The optimal noise disparity would select or not select 'l'. Now the value of optimal noise disparity is computed using a matrix structure in bottom to up approach. The array is computed using $N_D[l, d] = \max(N_D[l-1, d], n_l + N_D[l-1, d-d_l])$ row by row.

The first stage of dynamic programming in BSMV RGB-D is given by,

$$\begin{aligned} & \text{BSMV RGB-D}(v, d, n, n_1) \\ & \left\{ \begin{array}{l} \text{for } (d=0 \text{ to } n_1) N_D[0, d]=0 \text{ for } (l=1 \text{ to } n) \\ \quad \text{for } (d=0 \text{ to } n_1) \text{ of } (d[l] \leq d \\ N_D[l, d]=\max\{N_D[l-1, d], v[l]+N_D[l-1, d-d[l]]\} \\ \quad \text{else} \\ N_D[l, d]=N_D[l-1, d] \text{ return } N_D[n, n_1] \end{array} \right\} \quad (2) \end{aligned}$$

The algorithm is given below,

The next stage is to compute the suboptimal noise disparity. For this the auxiliary Boolean array $threshold[l, d]$ is 1, if the

l^{th} pixel in the $N_D[l, d]$ is selected, otherwise zero. Also, the optimal solution is to be formulated. The $threshold[l, d]$ is used for finding the suboptimal nose 'R' having maximum computation time. If $threshold[l, d]$ is 1 then $n \in R$, continuously repeating the process for $threshold[n-1, n_1-d_n]$. If $threshold[n, n_1]$ is 0 then $n \notin R$, continuously repeating the process for $threshold[n-1, n_1]$. The following partial code will compute the value of 'R'. Assigning n_1 to T for ($l = n$, down to 1) if $threshold[l, T] == 1$, output 1 and $T = T - d[l]$. The stage 2 of dynamic programming in BSMV RGB-D is given by $BSMV RGB-D(v, d, n, n_1)$, is shown at the top of the next page,

The algorithm is given below,

The application of DP completely removes the noise around the image border and the foreground object appears in better reconstructed form.

D. Pyramiding the Image

Pyramiding technique is used for the image with telescopic search to ease block matching. Given the full-scale image to detect the noise disparity, the process has to be carried over ± 7 -pixel block. By pyramiding, the image is downsized by a factor of 2. The process will reduce to ± 1 -pixel block. The noise disparity is then estimated on this down sized block. This down sized block act as the seed to the larger image. This enables search of the noise disparity only in smaller pixel block ranges and the process becomes at least five time faster compared to normal block matching. Here, three level image pyramiding is used.

Dynamic programming executed on noise disparity at every pyramidal level reduces the computational burden drastically. The depth map of the stereo image, the intrinsic parameters of the RGB-D camera and the concerned image are back projected on a pixel to pixel basis to form the 3D points. The

$$\text{intrinsic matrix I for RGB-D camera is, } I = \begin{bmatrix} f_L & S_L & C_L \\ 0 & f_R & C_R \\ 0 & 0 & 1 \end{bmatrix}$$

and the 3D image coordinates aligned with the homogenized stereo camera coordinates is $[L_c \ R_c]^T = I[L_c, R_c, Z_d]^T$. The intrinsic matrix describes all the pixels in the image which can be back projected into a 3D pixel. However, the distance of the concerned pixel to the camera is unknown. Considering S-B as the distance between the two RGB-D cameras, 'f' is the focal length, the noise disparity estimation of the depth map is given,

$$Z_d = \frac{f + D^{S-B}}{N_D} \quad (4)$$

The result of the re-projection shows the surrounding of the images appearing mutually orthogonal and the image is reconstructed accurately. Now performing a BSMV RGBD Gaussian N level pyramiding, a cascaded low pass filter produces down sampled images represented as $\vec{G}_p = [\vec{G}_0, \vec{G}_1, \vec{G}_2, \dots, \vec{G}_T]$. The separable N-dimensional kernel 't' = $[t_1 \ t_2 \ t_3 \ t_4 \ t_5]$ is used for construction of the down sampled images. In each direction a down sampling value of 2 is used.

$$\begin{aligned}
 & \text{BSMV RGB-D}(v, d, n, n_1) \\
 & \left. \begin{aligned}
 & \text{for } (d = 0 \text{ to } D) N_D[0, n_1] = 0; \text{ for } (l = 1 \text{ to } n) \text{ for } (d = 0 \text{ to } n_1) \\
 & \text{if } ((d[l] \leq d) \text{ and } (n[l] + N_D[l-1, d-d[l]] > N_D[l-1, d])) \\
 & \quad \{N_D[l, d] = n[l] + N_D[l-1, d-d[l]]; \text{Threshold}[l, T] = 1\} \\
 & \quad \text{else} \\
 & \quad \quad \{N_D[l, d] = N_D[l-1, d]; \text{Threshold}[l, T] = 0\} \\
 & \quad \quad T = n_1; \text{ for } (l = n \text{ down to } 1) \\
 & \text{If } (\text{Threshold}[l, T] == 1 \{ \text{output } l; T = T - d[l]; \} \text{ return } N_D[n, n_1] \}
 \end{aligned} \right\} \quad (3)
 \end{aligned}$$

Algorithm 1 To Calculate Optimal Noise Disparity

```

1: procedure BSMV RGB-D (v, d, n, n1)
2:   for d ← 0 to n1 ∧ ND[0, n1] == 0 do
3:     for l ← 1 to n do
4:       for d ← 0 to n1 do
5:         if (d[l] ≤ d) then
6:           ND[l,d] ← max{(ND[l-1,d], v[l] + ND[l-1,d-d[l]})}
7:         else
8:           ND[l,d] ← ND[l-1,d]
9:         end if
10:      end for
11:    end for
12:  end for
13:  return ND[n, n1]
14: end procedure

```

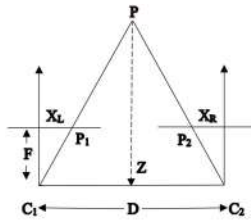


Fig. 2. Placement of RGB-D camera.

Mapping the N-dimensional kernel from one face to another is done as,

$$\vec{G}_{P+1} = M \vec{G}_P = \begin{bmatrix} 1 & 0 & 0 & \dots & \dots \\ 0 & 1 & 0 & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots \end{bmatrix} \begin{bmatrix} h & \dots & \dots \\ \vdots & h & \vdots \\ \dots & \dots & h \end{bmatrix} \vec{G}_P$$

Down sampling Convolution

IV. RESULTS AND DISCUSSION

This section presents the discussion on the results obtained by applying the proposed technique BSMV RGB-D to a 3D realistic head model and also to a fMRI image. Here the software used for simulation and testing is MATLAB R2018a [20] with neuroelectromagnetic forward head modelling toolbox (NFT) plug in. The 4×4 feature block matching technique has a higher rate of retrieval, as the shape of the focused image is retrieved well. The recovered image shows irregular depth range and spots of noise scattered across the image, especially

Algorithm 2 To Calculate Sub-Optimal Noise Disparity

```

1: procedure BSMV RGB-D (v, d, n, n1)
2:   for d ← 0 to n1 ∧ ND[0, n1] == 0 do
3:     for l ← 1 to n do
4:       for d ← 0 to n1 do
5:         if ((d[l] ≤ d) ∧ (n[l] +
           ND[l-1,d-d[l]] > ND[l-1,d])) then
6:           ND[l,d] ← n[l] + ND[l-1, d-d[l]]
7:           threshold[l,T] ← 1
8:         else
9:           ND[l,d] ← ND[l-1,d]
10:          threshold[l,T] ← 0
11:        end if
12:        T ← n1
13:        for d ← n to 1
14:          if (threshold[l,T]) == 1 then
15:            output l
16:            T ← T-d[l]
17:          end if
18:        end for
19:      end for
20:    end for
21:  end for
22:  return ND[n, n1]
23: end procedure

```

on the upper part of the image. There is excess noise scattering on the top of the image as there are no major features of the image available inside the 4×4 sub pixel matrix under comparison. The noise spot occurs because individual sub pixel disparity map is independent of the surrounding pixel.

The RGB-D cameras are kept approximately parallel while capturing multiple views of the image, to ensure that the disparity values are positive. Otherwise, the depth map will have both positive and negative values. The placement of the RGB-D cameras is illustrated in figure 2. Here, P1 and P2 are the projection of the cross-point P on the two images, C1 and C2 are the origin of RGB-D camera, F is the focal length, D is the distance between the two cameras and Z is the depth. The sub pixel estimation and correction are done by taking minimum mismatch of actual pixel and the neighboring pixel. A parabola is drawn considering these values and solved for minimum mismatch of the sub pixel. This results in the removal of the contouring effect and the estimation of disparity

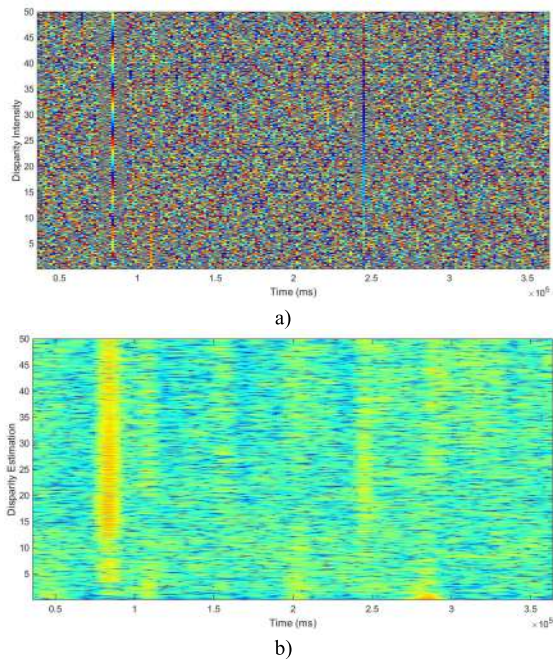


Fig. 3. a) Disparity intensity vs time b) disparity estimation vs time.

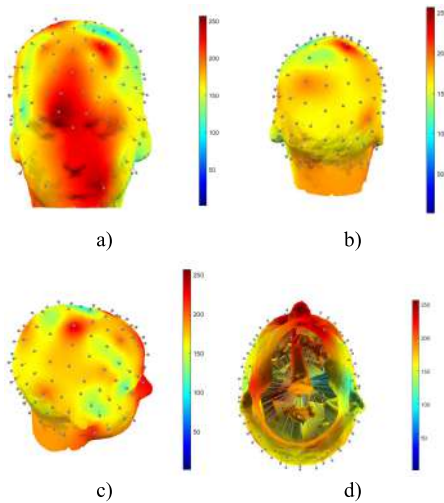


Fig. 4. IR variation in 16-bit RGB-D 3D realistic head model a) front view b) back view c) side view d) pixel mapped depth sensing view.

is fine-tuned. The result can be seen along the surface of the image.

Figure 3 shows the disparity intensity and disparity estimation of the image with varying time. The figure highlights the difference in the intensity of the pixels in the image captured by the RGB-D sensor.

A. BMSV RGB-D Applied on 3D Realistic Head Model

Figure 4 presents the IR variation in 16-bit RGB-D 3D realistic head model. These values are very vital for accurate disparity mapping. The front, back view and side views are obtained. Also, the pixel mapped depth sensing view is also presented. Figure 5 presents the IR variation in 64-bit RGB-D 3D realistic head model with various views. The significance

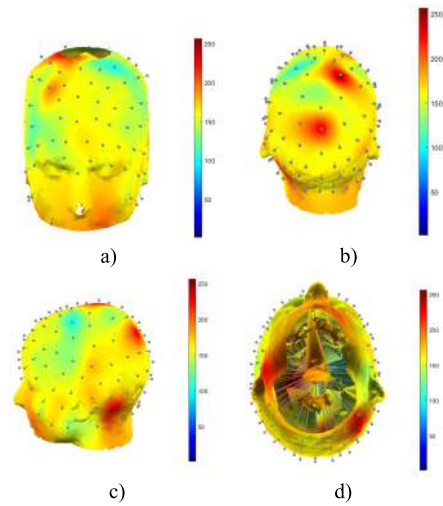


Fig. 5. IR variation in 64-bit RGB-D 3D realistic head model a) front view b) back view c) side view d) pixel mapped depth sensing view.

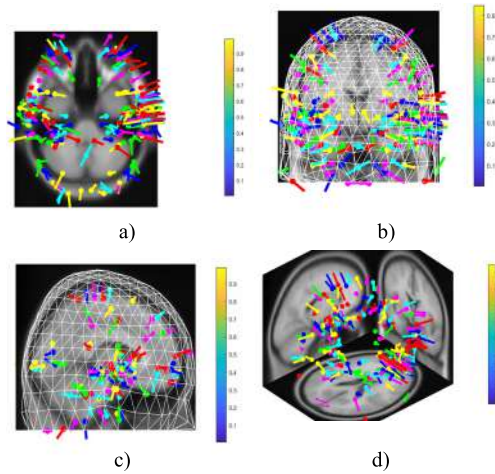


Fig. 6. 3D fMRI image a) top view b) sagittal view c) coronal view d) combined 3D view.

TABLE III
PSNR VARIATION FOR 8, 16 AND 32-bit SUB BLOCK FOR DIFFERENT BLOCK MATCHING METHOD

TSS (Three Step Search)	NTSS (New Three Step Search)	Four Step Search (FSS)	ARPS (Adaptive Rood Pattern Search)	BMSV RGB-D
35.52	36.44	36.87	37.79	39.81
35.17	35.55	36.39	37.12	38.09
35.01	35.50	36.10	37.02	38.17

of 16-bit and 64-bit is in the size of the physical memory support for RGB-D camera.

B. BMSV-RGBD applied on fMRI image

The top view, sagittal view and coronal view with mesh of 3D fMRI image is presented in figure 6. The weighted

TABLE IV
COMPARISON OF BSMV RGB-D WITH EXISTING CAMERA

Sensors (R) & Parameters (C)	Xtion Pro Live [21]	Orbbec Astra S [22]	Microsoft Kinect ii [23]	RS SR300 [24]	Intel D435 [25]	Proposed BSMV RGB-D
Technology	Infrared pattern	Structured light technology	Time of Flight (ToF)	Infrared pattern	Active stereoscopy	Infrared-Block Matching, Image Pyramiding and Dynamic Programming
Depth Range (m)	0.7 to 3.5	0.4 to 2	0.5 to 4.5	0.3 to 2	0.2 to 8	0.2 to 8
Field of View of Depth	58° H, 45° V, 70° D	60°H, 49.5°V, 73°D	70°H, 60°V	73° H, 59° V, 90° D	85.2 ° H, 58° V,	93.2° H, 68.5° V, 100.6° D
Frame rate (fps)	30, 60	30	30	30	30, 60, 90	30, 60, 90
Depth resolution (pixel)	628 × 468	640 × 480	512 × 424	640 x 480	1280×720	1280 × 720 pixel, 848 × 480, 640 × 480
Power consumption	< 2.5W	< 2.4W	~ 15W	1.8	600–1900 mW	600-1500 mW
RGB Resolution (pixel)	1280 × 1024	640 × 480	1920 × 1080	1920 × 1080	1920 × 1080	1920 × 1080, 1280 × 720, 848 × 480, 640 × 480

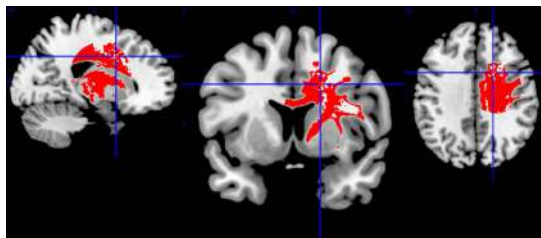


Fig. 7. Depth analysis of 3D fMRI image with BSMV-RGB-D.



Fig. 8. Image of an object with transparent surface (spectacles) captured and reconstructed using BSMV RGB-D method.

magnetic resonance images are analysed for cerebrospinal fluid, segmenting the skull, scalp and brain tissue. For the forward problem numerical solution, boundary element method (BEM) is used. The BEM meshes is generated once the segmented tissue volume is extracted. The individual realistic head model is wrapped by template head model electrode location if magnetic resonance image is not available. Here, the coordinates are aligned with the homogenized stereo camera. The intrinsic matrix describes all the pixels in the image which can be back projected into a 3D pixel. The result of the re-projection shows the surrounding of the images appears mutually orthogonal and the image is reconstructed accurately.

Figure 7 presents the depth analysis of 3D fMRI image with BSMV RGB-D. The cross point on the fMRI image indicates the seed pixel. The RGB-D image enabled with telescopic search is analysed using pyramiding technique for block matching. The noise disparity is detected for the full-scale image and it is applied over the 7-pixel block. Pyramiding will downsize the image by a factor of 2, which will further scale down to 1-pixel block. For this down sized block, the noise disparity is calculated. The down sized block will be the seed for the scaled up RGB-D image. The noise disparity is enabled for the search of the down sized smaller

pixel block. The search process is fivefold faster compared to existing block matching. The dynamic programming is run on the noise disparity which reduces the computational burden substantially. The IR projector emits the predefined pattern and the concerned RGB-D image is back projected on a pixel to pixel basis to form depth information.

The seed point in the image is identified initially and mapped to 3D space for the depth information. To identify the edge and mark it automatically, canny edge detection algorithm is used. Then the noise disparity is calculated from the cross mark on the two images. The placement of camera is shown in figure 2. The camera A is displaced to the right direction, that is in the X-direction positive D_x . The 3D coordinates of the left camera (X_1, Y_1, Z_1) would have ($X_1 - D_x, Y_1, Z_1$) coordinates in the right camera. This would project different image point X coordinates for the left and right camera creating noise disparity. The noise disparity value will be positive, negative or zero depending upon the position of the object compared to the cross point. This noise disparity is used to calculate the information of an object. The depth is then calculated by using the equation 4. The calibrated depth range obtained here is 0.2 m to 8 m for 16-bit and 64-bit realistic head models. Figure 8 presents the results obtained with the BSMV RGB-D with objects having transparent surface. Here,

the human face is sensed by the RGB-D camera and with the accurate depth information captured using the proposed method, the 3D image is reconstructed. The spectacle on the face is correctly decoded and shows a nullifying effect on the actual 3D object.

Table III presents the comparison of the proposed method with other existing block matching methods in terms of PSNR for varying sub block sizes. The proposed method BSMV RGB-D has better PSNR value compared to all the other existing block matching techniques and this signifies better performance. In block matching BMSV-RGBD, each pixel is having a very high correlation with its neighbors. This makes it easy to assign motion vector to a sub pixel block compared to individual pixels. In the BSMV-RGBD the RGB-D frame is segmented into $n \times n$ blocks. The current block of subsequent frame is aligned and matched with the corresponding subblock in the same coordinates of the previous frame for the pixel having a very large motion displacement. The optimized alignment and matching give the correct displacement.

V. CONCLUSIONS

This paper presents and discusses a novel system for object tracking and capturing of accurate depth information from shiny/transparent objects using RGB-D camera with rectified/non-rectified block matching and image pyramiding along with dynamic programming. The architecture and the working of the proposed system is presented and discussed in detail. Theoretical analysis of the BSMV dynamic programming algorithm is presented. The results achieved by the proposed method in simulations using MATLAB 2018a are presented. The results from the depth information analysis carried out on a 3D realistic head model and an fMRI image illustrate the accuracy of the proposed method. The advantages and better performance of the proposed method is also highlighted using a comparison with the existing RGB-D sensors and block matching techniques. Finally, the capture and accurate reconstruction of an object with transparent surface is demonstrated using the proposed method.

REFERENCES

- [1] D. Riahi, W. Bouachir, Y. Ouakrim, and N. Mezghani, "Depth imaging system for human posture recognition," in *Proc. IEEE 4th Middle East Conf. Biomed. Eng. (MECBME)*, Tunis, Tunisia, Mar. 2018, pp. 177–181.
- [2] C.-C. Sun, Y.-H. Wang, and M.-H. Sheu, "Fast motion object detection algorithm using complementary depth image on an RGB-D camera," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5728–5734, Sep. 2017.
- [3] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3443–3458, Aug. 2014, doi: 10.1109/tip.2014.2329776.
- [4] M. Gao, J. Jiang, G. Zou, V. John, and Z. Liu, "RGB-D-based object recognition using multimodal convolutional neural networks: A survey," *IEEE Access*, vol. 7, pp. 43110–43136, 2019.
- [5] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments," *Int. J. Robot. Res.*, vol. 31, no. 5, pp. 647–663, Apr. 2012.
- [6] Z. Cai, Y. Long, and L. Shao, "Adaptive RGB image recognition by visual-depth embedding," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2471–2483, May 2018.
- [7] C. Jing, J. Potgieter, F. Noble, and R. Wang, "A comparison and analysis of RGB-D cameras' depth performance for robotics application," in *Proc. 24th Int. Conf. Mechatronics Mach. Vis. Pract. (M2VIP)*, Auckland, New Zealand, Nov. 2017, pp. 1–6.
- [8] Y. W. Kuan, N. O. Ee, and L. S. Wei, "Comparative study of intel R200, Kinect v2, and primesense RGB-D sensors performance outdoors," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8741–8750, Oct. 2019.
- [9] P. S. Santoso and H.-M. Hang, "Learning-based human detection applied to RGB-D images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Beijing, China, Sep. 2017, pp. 3365–3369.
- [10] D. K. Jain, S. Jacob, J. Alzubi, and V. Menon, "An efficient and adaptable multimedia system for converting PAL to VGA in real-time video processing," *J. Real-Time Image Process.*, pp. 1–13, 2019, doi: 10.1007/s11554-019-00889-4.
- [11] T.-H. Chien, Y.-H. Lu, J.-W. Lin, and R.-G. Chang, "An enhanced depth estimation system using RGB-D cameras and gyroscopes," in *Proc. Int. Conf. Appl. Syst. Innov. (ICASI)*, Sapporo, Japan, May 2017, pp. 1554–1557, doi: 10.1109/ICASI.2017.7988224.
- [12] M. Goffredo et al., "Shape analysis of bicipital contraction by means of RGB-D sensor, parallel transport and trajectory analysis," in *Proc. XIV Medit. Conf. Med. Biol. Eng. Comput.*, Cyprus, Middle East, 2016, pp. 634–639.
- [13] H. Wang, Y. Tian, and Y. Zhang, "A novel depth propagation algorithm with color guided motion estimation," in *Proc. 2013 Vis. Commun. Image Process. (VCIP)*, Kuching, Malaysia, Nov. 2013, pp. 1–5, doi: 10.1109/VCIP.2013.6706419.
- [14] F. Liu, "3D block matching algorithm in concealed image recognition and E-commerce customer segmentation," *IEEE Sensors J.*, to be published.
- [15] A. A. Altahir et al., "Optimizing visual surveillance sensor coverage using dynamic programming," *IEEE Sensors J.*, vol. 17, no. 11, pp. 3398–3405, Jun. 2017.
- [16] P. G. Vinoj, S. Jacob, V. G. Menon, S. Rajesh, and M. R. Khosravi, "Brain-controlled adaptive lower limb exoskeleton for rehabilitation of post-stroke paralyzed," *IEEE Access*, vol. 7, pp. 132628–132648, 2019, doi: 10.1109/ACCESS.2019.2921375.
- [17] S. Rajesh, V. Paul, V. G. Menon, S. Jacob, and P. Vinod, "Secure brain to brain communication with edge computing for assisting post-stroke paralyzed patients," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2019.2951405.
- [18] S. Jacob, V. G. Menon, F. Al-Turjman, V. P. G., and L. Mostarda, "Artificial muscle intelligence system with deep learning for post-stroke assistance and rehabilitation," *IEEE Access*, vol. 7, pp. 133463–133473, 2019.
- [19] V. G. Menon, S. Jacob, S. Joseph, and A. O. Almagrabi, "SDN powered humanoid with edge computing for assisting paralyzed patients," *IEEE Internet Things J.*, to be published, doi: 10.1109/JIOT.2019.2963288.
- [20] Accessed: Apr. 10, 2019. [Online]. Available: https://in.mathworks.com/products/new_products/release2018a.html
- [21] Accessed: Apr. 17, 2019. [Online]. Available: https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/specifications/
- [22] Accessed: Apr. 17, 2019. [Online]. Available: <https://orb3d.com/product-astra-pro/>
- [23] E. Lachat, H. Macher, T. Landes, and P. Grussenmeyer, "Assessment and calibration of a RGB-D camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling," *Remote Sens.*, vol. 7, no. 10, pp. 13070–13097, Oct. 2015.
- [24] Accessed: Apr. 17, 2019. [Online]. Available: <https://ark.intel.com/content/www/us/en/ark/products/92329/intel-realsense-camera-sr300.html>
- [25] Accessed: Apr. 17, 2019. [Online]. Available: <https://ark.intel.com/content/www/us/en/ark/products/128255/intel-realsense-depth-camera-d435.html>



Sunil Jacob (Member, IEEE) received the Ph.D. degree in electronics and communication engineering from Bharathiar University, India, in 2015. He is currently the Director with the Center for Robotics, and also a Professor with the Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, India. His research interests include SDN, edge computing, brain computer interface, and signal processing.



Varun G. Menon (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from Sathyabama University, India, in 2017. He is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include sensors, IoT, fog computing, and underwater acoustic sensor networks. He is currently serving as the Guest Editor for the IEEE SENSORS JOURNAL, the *IEEE Internet of Things Magazine*, and the IEEE TRANSACTIONS ON INDUSTRY INFORMATICS. He is also a Distinguished Speaker of ACM.



Saira Joseph (Member, IEEE) received the M.Tech. and Ph.D. degrees in electronics and communication engineering from the Cochin University of Science and Technology, Kerala, India, in 2006 and 2017, respectively. She is currently an Associate Professor and the Head of Department of Electronics and Communication Engineering, SCMS School of Engineering and Technology, Kerala, India. Her interests include sensors, edge computing, fractal antennas, UWB, RADAR, and metamaterials.

[Log in](#)[Register](#)[Cart](#)[IETE Technical Review >](#)

Volume 38, 2021 - Issue 4

854 | 23

Views | CrossRef citations to date | 0 | Altmetric

Articles

Blockchain For Intelligent Transport System

Anandkumar Balasubramaniam, Malik Junaid Jami Gul , Varun G. Menon & Anand Paul 

Pages 438-449 | Published online: 31 May 2020

 [Download citation](#)  <https://doi.org/10.1080/02564602.2020.1766385> [Check for updates](#) [Full Article](#) [Figures & data](#) [References](#) [Citations](#) [Metrics](#) [Reprints & Permissions](#)[Read this article](#)

ABSTRACT

Intelligent Transportation System (ITS) is gaining attention but at the same time, road accidents, congestion, delays, etc. have also increased. Relative information about such events is vital. Such information can be presented in legal processes as digital proof. Availability of the information is not a problem as multidimensional data have been recorded all the time by ITS. Recording all the information in ITS arises the problem of fetching relevant information and removing other facts and figure that are not required to describe certain situations such as an accident. To

address this issue, we analyze road accident data and reduce various dimensions with Principal Component Analysis (PCA). Linear Discriminant Analysis (LDA) and

[Home](#) ▶ [All Journals](#) ▶ [IETE Technical Review](#) ▶ [List of Issues](#) ▶ [Volume 38, Issue 4](#)
▶ [Blockchain For Intelligent Transport Sys ...](#)

three datasets where error rate for PCA is 32% with Dataset1. Likewise, error rate for LDA and NMF are 36% and 35%, respectively. While keeping in mind that such reduced data is helpful in many legal processes, we introduce Blockchain in the framework. Blockchain can make data immutable thus can be considered as digital proof. Blockchain also requires a smart contract in this situation between insurance companies to collect data in case of any uncertain situation. Such analysis can offer a different point of views and trends in data. Information can be more explainable to define the situation and helps to develop a friendly environment for day-to-day customers. The proposed framework provides dimensionality reduction of data that eventually reduce the data dimension to store in Blockchain.

Q KEYWORDS: Accident analysis Blockchain ITS Principal component analysis
Smart transportation system Traffic surveillance

Acknowledgements

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF2017R1C1B5017464).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Additional information

Funding

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government [grant number 2020R1A2C1012196].

Notes on contributors



Anandkumar Balasubramaniam

Anandkumar Balasubramaniam received his bachelors in electronics and communication engineering under Anna University, Chennai, India in 2012. Currently, he is pursuing his masters combined PhD program under the supervision of Prof Anand Paul at Kyungpook National University, Daegu, South Korea. His research areas include intelligent transportation systems, internet of vehicles, smart city and smart vehicles. **Email:** bsanandkumar@gmail.com



Malik Junaid Jami Gul

Malik Junaid Jami Gul received his master's degree in computer science and information technology, Pakistan in 2015. Currently, he is pursuing his PhD with Dr Anand Paul at Kyungpook National University, Daegu, South Korea. His research interests include big data analytics, internet of things, smart systems, network traffic analysis and monitoring, smart city, operating system security, intrusion detection, and computer and network security. **Email:** junaidgul@live.com.pk



Varun G. Menon

Varun G Menon (SM'19) is currently an associate professor with the

[Home](#) ▶ [All Journals](#) ▶ [IETE Technical Review](#) ▶ [List of Issues](#) ▶ [Volume 38, Issue 4](#)

▶ [Blockchain For Intelligent Transport Sys ...](#)



Engineering and Technology, India. He has completed his PhD in computer science and engineering from Sathyabama University, India in 2017. His research interests include sensors, IoT, fog computing, brain computer interface and underwater acoustic sensor networks. He is a distinguished speaker of ACM. He is currently associate editor for *IET Quantum Communications Journal*, Editorial Member of *IEEE Technology Policy and Ethics* and a guest editor for *IEEE Sensors Journal*, *IEEE Internet of Things Magazine*, *IEEE Transactions on Industry Informatics*, *Computer Communications Journal* and the *Journal of Supercomputing*. **Email:** varunmenon@ieee.org



Anand Paul

Anand Paul received the PhD degree in electrical engineering from the National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently working as an associate professor in the School of Computer Science and Engineering, Kyungpook National University, South Korea. He is a delegate representing South Korea for M2M focus group and for MPEG. His research interests include algorithm and architecture reconfigurable embedded computing. He is IEEE Senior member and has guest edited various international journals and he is also part of Editorial Team for *Journal of Platform Technology*, *ACM Applied Computing review* and *Cyber-Physical Systems*. He serves as a reviewer for various IEEE /IET/Springer and Elsevier journals. He is the track chair for Smart human computer interaction in ACM SAC 2015, 2014. He was the recipient of the Outstanding International Student Scholarship Award in 2004–2010, the Best Paper Award in National Computer Symposium and in 2009, and organized an International Conference on Soft computing and Network Security, India in 2015.

Related research

People also read

Recommended articles

Cited by

[Home](#) ▶ [All Journals](#) ▶ [IETE Technical Review](#) ▶ [List of Issues](#) ▶ [Volume 38, Issue 4](#)
▶ [Blockchain For Intelligent Transport Sys ...](#)

[Blockchain in transport and logistics – paradigms and transitions](#) >

Lenny Koh et al.

International Journal of Production Research

Published online: 6 Apr 2020

[Blockchain applications in supply chains, transport and logistics: a systematic review of the literature](#) >

Mehrdokht Pournader et al.

International Journal of Production Research

Published online: 11 Aug 2019

[A comprehensive view of intelligent transport systems for urban smart mobility](#) >

Riccardo Mangiaracina et al.

International Journal of Logistics Research and Applications

Published online: 13 Oct 2016

[View more](#)

[Home](#) ▶ [All Journals](#) ▶ [IETE Technical Review](#) ▶ [List of Issues](#) ▶ [Volume 38, Issue 4](#)
▶ [Blockchain For Intelligent Transport Sys ...](#)

[Home](#) ▶ [All Journals](#) ▶ [IETE Technical Review](#) ▶ [List of Issues](#) ▶ [Volume 38, Issue 4](#)
▶ [Blockchain For Intelligent Transport Sys ...](#)

[Home](#) ▶ [All Journals](#) ▶ [IETE Technical Review](#) ▶ [List of Issues](#) ▶ [Volume 38, Issue 4](#)

▶ [Blockchain For Intelligent Transport Sys ...](#)

[Authors](#)

[R&D professionals](#)

[Editors](#)

[Librarians](#)

[Societies](#)

[Opportunities](#)

[Reprints and e-prints](#)

[Advertising solutions](#)

[Accelerated publication](#)

[Corporate access solutions](#)

[Overview](#)

[Open journals](#)

[Open Select](#)

[Dove Medical Press](#)

[F1000Research](#)

[Help and information](#)

[Help and contact](#)

[Newsroom](#)

[All journals](#)

[Books](#)

Keep up to date

Register to receive personalised research and resources by email

 [Sign me up](#)

[Copyright © 2023 Informa UK Limited](#) [Privacy policy](#) [Cookies](#) [Terms & conditions](#) [Accessibility](#)



Registered in England & Wales No. 3099067
5 Howick Place | London | SW1P 1WG

Dogo Rangsang Research Journal
(A Bilingual Research Journal, Indexed in UGC-Care list)
Vol-12, Issue. 1 No.02 January 2022

ISSN : 2347-7180

DOGO RANGSANG

Research Journal

দগো ঝাংছাং

গবেষণা পত্রিকা



CHIEF EDITOR (HON.):
Dr. Upen Rabha Hakacham
EDITORS (HON.):
Dr. Lalit Chandra Rabha
Dr. Neeva Rani Phukan

মুখ্য সম্পাদক (অবৈতনিক):
ড° উপেন ঝাভা হাকাচাম
সম্পাদকদ্বয় (অবৈতনিক):
ড° ললিত চন্দ্র ঝাভা
ড° নিভা ঝাগী ফুকন

A Peer Reviewed Bilingual Research Journal
(Indexed in UGC-CARE List)

ISSN 2347-7180

DOGO RANGSANG RESEARCH JOURNAL
দগো বাংছাং গৱেষণা পত্ৰিকা

Chief Editor (Hon.) : Dr. Upen Rabha Hakacham
Editors (Hon.) : Dr. Lalit Chandra Rabha
Dr. Neeva Rani Phukan

মুখা সম্পাদক (অবৈতনিক) : ড° উপেন ৰাভা হাকাচাম
সম্পাদকসমূহ (অবৈতনিক) : ড° ললিত চন্দ্ৰ ৰাভা
ড° নিজা ৰাণী ফুকন



Dogo Rangsang Research Society
Reg. No. KAM-M/263/L/ 595 of 2015-16
Gauhati University Campus
Guwahati - 781014

Dogo Rangsang Research Journal (ISSN : 2347-7180)
(A Bilingual Research Journal of Social Science and Humanities indexed in UGC-CARE List.)

EDITORIAL BOARD :

Advisers :

- Dr. Biplab Chakravarty, Retired Professor, Dept. of Bengali, Vardhaman University.
Dr. K. V. Subbarao, Retired Professor, Dept. of Linguistics, Delhi University.
Dr. Prabin Ch. Das, Retired Professor, Dept. of Folklore, Gauhati University.
Dr. Irshad Ali, Retired Professor, Dept. of Anthropology, Gauhati University.
Dr. Dipti Phukan Patgiri, Prof. and HOD, Dept. of Assamese, Gauhati University.

Reviewers of Papers :

1. Dr. Ajit Kumar Baishya, Professor, Department of Linguistic, Assam University, Silchar.
2. Dr. Nava Kr. Handique, Professor, Department of Assamese, Dibrugarh University, Dibrugarh.
3. Dr. Dilip Kalita, Director, ABILAC, Guwahati.
4. Dr. Dipak Kr. Roy, Professor, Department of Bengali, Raiganj University, West Bengal.
5. Dr. Jyotirekha Hazarika, Associate Professor, Department of Assamese, J.B. College (Autonomous), Jorhat.
6. Dr. Prafulla Kr. Nath, Professor, Department of Assamese, Guahati University.
7. Dr. Sangeeta Saikia, Professor, Viswa Bharati Niketan.
8. Dr. Rabindra Nath Sarma, Professor & Dean, Jharkhand University, Jharkhand.
9. Dr. Sumi Kalita, Assistant Professor, Department of Assamese, Bodoland University, Kokrajhar.

Internal Reviewers of this Issue :

1. Dr. Upen Rabha Hakacham (Chief Editor)
2. Dr. Lalit Ch. Rabha (Honorary Editor)
3. Dr. Neeva Rani Phukan (Honorary Editor)

Chief Editor (Hon.) :

Dr. Upen Rabha Hakacham
Professor and Former Head, Dept. of Assamese, Gauhati University.

Editors (Hon.) :

Dr. Lalit Chandra Rabha, Principal, Dudhnoi College, Dudhnoi.
Dr. Neeva Rani Phukan, Associate Professor, Assamese, KKHSOU, Guwahati-17

Published by Dr. Angshuman Das, Secretary,
Dogo Rangsang Research Society, Gauhati University Campus, Guwahati-14
and Printed at Dream Graphics, Naokata, Baksa (BTAD), Assam,
E-mail : editor.drsjournal@gmail.com

INDEX

S.No	TITLE	Page No
1	CUSTOMER SATISFACTION ON E BANKING SERVICES - A COMPARATIVE STUDY WITH REFERENCE TO INDIAN BANK AND AXIS BANK IN CHENNAI CITY	1
2	CREATION OF THE WORLD ... EMERGENCE OF GALAXY... FINAL SOLUTION TO BIG-BANG... JAMES WEBB SPACE TELESCOPE	7
3	GEORGE ORWELL'S CONTRIBUTION IN ENGLISH LITERATURE	10
4	CSR- CORPORATE SOCIAL RESPONSIBILITY	14
5	LOOK EAST TO ACT EAST POLICY: NATIONAL PRINT MEDIA FRAMING OF THE POLICY	17
6	A STUDY ON THE ADJUSTMENT OF SECONDARY SCHOOL STUDENTS OF DHANBAD DISTRICT, JHARKHAND	26
7	FROM LOOK EAST TO ACT EAST – AN ANALYSIS OF INDIA'S EXISTING POLICY PARADIGM AND ITS IMPACT ON NORTH EAST INDIA	29
8	ARTICULATING BHANUMATI AS A 'STRONG WOMAN': A STUDY OF THE FIRST ASSAMESE NOVEL <i>BHANUMATI</i>	38
9	PREPARATION OF LICE KILLER OIL FROM ILLICIMUM VERUM, SYZYGIIUM AROMATICUM & CINNAMOMUM VERUM	43
10	CYBER SECURITY BEHAVIOUR AMONG THE COLLEGE STUDENTS IN COIMBATORE DISTRICT	48
11	PREPARATIVE AND QUALITATIVE ANALYSIS OF PROTEIN POWDER	55
12	BENEFITS OF DIGITAL MARKETING IN THE RETAIL SECTOR	61
13	FACTORS AFFECTING ONLINE SHOPPING IN THANJAVUR DISTRICT	66
14	<u>COALESCING BLENDED LEARNING AND FLIPPED CLASSROOMS</u>	73
15	BIODIESEL PRODUCTION FROM CANOLA OIL WITH KOH BY RANSESTERIFICATION PROCESS	78
16	PREPARATION AND EVALUATION OF BIOTIN POWDER	83

COALESCING BLENDED LEARNING AND FLIPPED CLASSROOMS

Divya M S, Assistant Professor, SCMS School of Engineering & Technology, Vidya Nagar, Palissery, Karukutty, Ernakulam – 683 576, Kerala (Affiliated to APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala)

Abstract

Learning and teaching process takes a new track detouring itself from the conventional mode. Applying technology with proper pedagogy and planning can create wonders. Major takeaway of Flipped classroom learning is, its student centered and creates a self-paced environment. Blended learning is a pedagogical approach which combines the traditional teaching methods with E-learning technology. Technology works collaboratively with a conventional mode of learning and teaching. Complete potential of teachers and students can be better utilized in such a way of teaching and learning. Both teachers and students are upgraded to a next higher level wherein novel ways of teaching and learning can be enjoyed. Learning can be made more fun than being stressed. National Education Policy, 2020 have initiated a major shift in the teaching and learning process. Multidisciplinary education can ensure a promising future as it inculcates a vast diverse area of knowledge. A new era of teaching and learning aided with technology can definitely strive our nation towards progress. The paper focuses on the use of technology in the flipped classroom and its implementations and challenges.

Key words: Flipped classrooms, Blended learning, E-learning, National Education Policy

Introduction

The major takeaway of this pandemic is that we began to think and make happen the impossible and unknown techniques of teaching and learning. Coalescence of blended learning and flipped classrooms is the need of the era. Blended learning creates a more efficient and effective way of learning and teaching. Educational sectors began to follow the work from home approach, inculcating both the techniques. Slowly we began to upgrade ourselves to the latest technology and being equipped, makes it more effective and accessible.

Statement of the Problem

Reformations are time taking and this pandemic had paved way for innovations in the teaching and learning process. The coalescence of flipped classroom and blended learning have proved to be more effective and beneficial to both teachers and students. This study may create an awareness among various educational institutions regarding the amalgamation of both the system of education. This study has made an exploratory investigation among instructors and students with a questionnaire to get more insight into the respondent's opinion. This study has identified the convenience, accessibility and technological knowledge of the users.

Objectives of the study

- To identify the respondent's view in flipped classroom learning
- To analyze the reason behind why are why doesn't they use flipped mode
- To analyze the impact of both learning approach and to find out which is effective

Review of Literature

Renwick, Matt (2016) had identified the emerging importance of blended and flipped mode of learning during the pandemic and post pandemic era. It's the need of the hour to upgrade and update ourselves to the latest and most effective techniques of education. Both the learning techniques have its own pros and cons.

Linton, Jayme (2008) studied the factors affecting the respondents view about the videos posted as a part of flipped classroom learning. These videos could enrich their knowledge level as they are imbibing the subject prior to the classroom lectures. This gives the learners additional information

about their subject and would provoke innovative thoughts and interest in them, which could be further developed through classroom discussions.

Mayer RE, ed. (2003) focused on the updated subject knowledge comprehended through interactive sessions in blended classrooms. Traditional way of learning limits their upgradation of knowledge or it would enhance more information only to those with research skills. Whereas, flipped learning enables equal opportunity to all the learners without any categorization. Linton, Jayme (2008) focused on the future scope of teaching and learning to integrate a holistic approach through multidisciplinary and integrated learning.

Research Methodology

This study has used deductive design. The respondent view is taken into consideration in the study area. The convenient sampling method is used for this research work. This study considers the respondents using both flipped and blended classroom learning. Life Skills subject learners were taken into consideration. The online websites, videos and study materials were referred for secondary data collection. The questionnaire consists of more than fifteen questions. The Likert five point scales are used to collect the respondent's opinion.

Results and Discussion

A survey is taken about the respondents as a part of Flipped classroom learning through google classroom as platform.

Table No. 1: Blended learning analysis

Flipped video topic and key concepts	Active learning strategy
Week 1: Life Skills	In small groups, students construct a mind map that illustrates the key ideas associated with the ten core life skills identified by WHO and the contributing factors.
Week 2: Self Awareness	In doing so, students consider the following questions: How is the world's behavior trending? Why? Drawing on Reading 1 (Renwick, Matt (2016), what are the strength and weakness? Go for a SWOT analysis.
Week 3: Morals, values and Ethics	Students complete the What I Know and What I Want to Know section of the chart and return it to their lecturer
Week 4: Stress Management	Identifying the stress management techniques and methods to make it effective.
Week 5: Critical & Creativity thinking	Brainstorm the points to be highlighted in each and identifying the six thinking hats.
Week 6: Interview Skills	Mock Interviews are being done to update themselves with the possible questions that can be asked.
Week 7: Communication	In pairs, students draw labelled diagram (mind map) that illustrates the leadership skills and compare, discuss and refine diagrams with a neighboring pair

Per week each tutorial videos have been uploaded in Google classroom and the number of respondent has been noted. An initial introduction has been given which will aid them to get an overall idea about the subject matter. Detailing will help them to prepare ahead for the upcoming sessions.

Table No. 2: Respondents view on Flipped classroom

Flipped Videos	Viewed by n (%) respondents	Number of times respondents viewed the videos (n)					Mean views (n)
		0	1	2	3	> 4	
Week 1: Life Skills	163 (95.3)	25	118	15	1	4	2.1
Week 2: Self Awareness	160 (93.6)	25	105	25	0	5	2.2

Week 3: Moral, Values & Ethics	158 (92.4)	21	83	41	8	5	2.5
Week 4: Stress Management	159 (93.0)	33	71	42	6	7	2.4
Week 5: Critical & Creative thinking	158 (92.4)	30	92	26	6	4	2.2
Week 6: Interview Skills	153 (89.5)	25	78	40	2	8	2.4
Week 7: Communication	151 (88.3)	30	93	21	1	6	2.2
I did not watch any of the videos	10 (5.8)						

A summary of the flipped videos viewed by survey respondents during the semester, and the number of times they were viewed (n = 171). As a part of Flipped Classroom learning, a survey was taken to analyze the respondents about the study videos posted. This study selects certain topics and assign them weekly, calculation is done by considering the respondents view and number of times they have viewed the study material. Flipped classroom learning expects a learner to preview the assigned works and come prepared for the discussion class. Number of weeks have been taken into consideration and percentage of viewers to the videos and number of times the videos have been repeatedly seen is take into consideration to calculate the Mean viewers. This gives us an idea of the number of students who are taking the flipped mode seriously and responding. The number of repeated views indicate the importance and relevance of topics and the quality of lectures posted. (250)

Table No. 3: A summary of when students viewed the flipped videos (n = 161)

Response options	Responses n (%)
Before lectures	128 (79.5)
During lectures	6 (3.7)
After lectures	64 (39.8)
Other	12 (7.5)

Table No. 4: A summary of students reasons why they didn't watch the flipped videos

Reason	Frequency
Lack of time	32
Forgot	22
Did not need to watch them	5
Problem with Internet access or technical issues	5
Lack of personal motivation or interest	4
Read the flipped video notes instead	2
Lack of personal organization	2

Table No. 5: A summary of the ways in which the flipped videos were most helpful to students learning

Themes	Frequency
Provided requisite knowledge for lectures and/ or tutorials	62
Clear and concise explanations of key concepts/ content	39
Supported understanding of key concepts	28
Visual representation of information	12
In-depth explanations of key concepts	12
Introduced new topics/ key concepts	11
To consolidate and revise learning	10
Ability to pause and replay videos	10
Fostered deep learning	6

Auditory representation of information	4
An additional resource for learning	3
The videos were enjoyable/ engaging	2
Provision of notes to accompany the videos	2
Flipped classroom learning had found out to be most effective as indicated by the tables given above.	
More than majority of the learners found it to be the most effective way of achieving learner outcomes.	

Table No. 6: Questionnaire 1

Sr. no	Questions	Suggestions
1	Please indicate which of the following seven flipped videos you watched in Google classroom and estimate the number of times you watched each one.	
2	During final year exam preparation, will revisit the videos.	
3	When do you usually watch the videos? Before Lectures During Lectures After Lectures	YES / NO
4	Do you feel the videos helped you to understand the module content? a. If YES: in what way were the videos most helpful to your learning? b. If NO: Why were they not helpful?	YES / NO
5	In what ways, if any, could the videos be improved?	

Table No. 6: Questionnaire 2

Sr. no	Questions	Suggestions
1	Have you watched all the videos	YES / NO
2	If NO: state the reason	YES / NO
3	Why didn't you watch them?	YES / NO
4	Rate each item to indicate your agreement (strongly agree, agree, neither agree/disagree, disagree, strongly disagree) a. The flipped classroom approach made me feel more motivated to learn the content within the module. b. More engaging and informative than traditional learning. c. Would not recommend the flipped classroom approach. d. Would recommend for traditional classroom approach than flipped classroom	YES / NO YES / NO YES / NO YES / NO

Conclusion


We envision for a greater change to happen in our educational sectors by the holistic and multidisciplinary approach in the upcoming New Education Policy 2020. The multidimensional aspect of a student and interdisciplinary way of education can enlighten our young minds to excel them in future. We need a blending of such education to upgrade the quality of wisdom that is being imparted. There existed a digital divide within the educators and students on using the latest technology, but with this pandemic and increasing usage of technology we got adapted to it. Now we began to equip ourselves with the latest and updated technologies that are available. The pandemic had up sided our economy to a greater extend and we are still struggling to find a way out for existence. But it had also opened way for more solutions which we thought would never happen. When we began to accept changes and move according it to, we began to cross our limits and conquer heights. Never ever we thought for such a drastic change to happen in educational sectors that would create history. Initial

process to adjust and update ourselves was a bit struggle but with hands together, we could overcome all the obstacles and move ahead. Let's envision for a much better and bright future ahead us and we can be the torch bearers for it.

References

1. Arney, Liz Go blended! : a handbook for blending technology in schools San Francisco, CA: Jossey-Bass, 2015.
2. Bermann J, Sans A. Flip Your Classroom: Reach Every Student in Ebery Class Everyday. Washington, DC: ISTE; and Alexandria, VA: 2012.
3. Fulton, Kathleen P. Time for learning : top 10 reasons why flipping the classroom can change education Thousand Oaks, CA: Corwin, 2014.
4. Goodwin B, Miller K. Evidence on flipped classrooms is still coming in. Edu Leadersh. 2013;70:78-79
5. Linton, Jayme The blended learning blueprint for elementary teachers Thousand Oaks, CA: Corwin, 2018.
6. Mayer RE, ed. The Cambridge Handbook of Multimedia Learning. New York, NY: Cambridge University Press; 2005.
7. McLaughlin JE, Roth MT, Glatt DM, et al. The flipped classroom: a course redesign to foster learning and engagement in a health professions schools. Acad Med. 2014
8. Renwick, Matt 5 myths about classroom technology : how do we integrate digital tools to truly enhance learning? Alexandria, VA: ASCD, 2016.
9. Sheninger, Eric C. Uncommon learning : creating schools that work for kids Thousand Oaks, CA: Corwin, 2016.
10. Tucker, Catlin R.; Wycoff, Tiffany Blended learning in action : a practical guide toward sustainable change Thousand Oaks, CA: Corwin, 2017.

ISSN 0976 - 8165



THE CRITERION


AN INTERNATIONAL JOURNAL IN ENGLISH

11th Year of Open Access


**Bi-Monthly Refereed and Peer-Reviewed
Open Access e-Journal**

Vol. XI, Issue-3 (June 2020)

Editor-In-Chief : Dr. Vishwanath Bite
Managing Editor : Dr. Madhuri Bite



www.the-criterion.com



AboutUs: <http://www.the-criterion.com/about/>

Archive: <http://www.the-criterion.com/archive/>

ContactUs: <http://www.the-criterion.com/contact/>

EditorialBoard: <http://www.the-criterion.com/editorial-board/>

Submission: <http://www.the-criterion.com/submission/>

FAQ: <http://www.the-criterion.com/fa/>



ISSN 2278-9529

Galaxy: International Multidisciplinary Research Journal
www.galaxyimrj.com

Exploring the Narratives of Human Resilience in History and Highlighting their Significance in Present Times as in Anne Frank's *The Diary of a Young Girl*

Divya MS

Assistant Professor,

Department of English,

SCMS School of Engineering and Technology, Ernakulam.

Article History: Submitted-22/05/2020, Revised-30/06/2020, Accepted-01/07/2020, Published-10/07/2020.

Abstract:

“We realize the importance of our voice only when we are silenced”

Remembering the quotes of Malala Yousafzai, I would like to travel through the memoirs of holocaust faced by the Jews during Hitler's reign as marked in the writings of Anne Frank in her famous work “The Diary of a Young Girl”. It is indeed true that throughout the history, hundreds of thousands of individuals have undergone heart rending suffering and horrors beyond their worst dreams. Humans have, time and again exhibited extraordinary resilience in adapting to the situation. During these lock down days we do face a lot of discomfort and frustration to be confined into our own safe home. But just think about the thousands and millions who had gone on exile on fear of death and about the cruelties they had undergone. My attempt here is to pen down the agonies and resilience they had faced during these holocausts. Anne, a young Jewish girl is forced into hiding with her family and one other family in Nazi occupied Amsterdam. The inscriptions in the form of diary writing tells us about her feelings and experiences they had faced and also about her budding hopes to be free once again. Things began to change when the Nazis came to power. Their aim was to remove the Jews from German society even though they were less than 1% of the population. Nazi believed that Jews were the root of all the evils. Life was horrific for Jews and they began to flee from Germany. Nazis burned down the synagogues and Jewish owned shops and even burned their books. Jews were fleeing and tried to find shelter wherever they could. Nazis deported these people to forced labour camps, where they worked to produce supplies for the increasingly strained war economy. In most camps the prisoners were devoid of sufficient food, equipment, medicine and clothing. There was a complete disregard and their health was deteriorating day by day. As a result of these conditions, death rates in labour camps were extremely high.

Believing Holland was safe for Jews, Anne's family moved to Amsterdam in 1933. 'The Diary of a Young Girl' also known as 'The Diary of Anne Frank', a book of diary writings kept by Anne Frank while she was hiding for two years in the secret annex, with her family during the Nazi occupation of the Netherlands. The family was apprehended in 1944 and Anne Frank died of typhus in the Bergen-Belsen concentration camp in 1945. The diary was retrieved by Anne's father Mr. Otto Frank, the family's only known survivor after the war. The writings were from 14th June 1942 to 1st August 1944. Her father gifted her a red checked diary on her 13th birthday, June 12th 1942. It was not like a usual diary writing, she wrote as letters to her best friend that is, diary whom she addressed as Kitty. In August 1944, they were caught from the secret annex and were deported to Nazi concentration camps. Anne died when she was just fifteen years old. These letters were not just the experiences of a thirteen-year-old young girl, it gives us an insight into the most terrific inhumane situation that mankind had ever undergone. The wordings which she breathed became eternal and true.

"I want to go on living even after my death"

Keywords: Holocaust, Concentration Camps, Resilience.

"I want to bring out all kinds of things that lie buried deep in my heart".

As rightly said by Anne Frank, this is exactly what her writings were. A mixture of agonies, frustration, happiness, fear, realisations, relations, her first love and more over an account of what happened in the outside world. These letters were not just the experiences of a thirteen-year-old young girl, it gives us an insight into the most terrific inhumane situation that mankind had ever undergone. An account for her journals do tell us about how much Jews had suffered and deprived from the rest of the society. The restrictions imposed on them were even more harsh. They must always wear a yellow star and had to be indoors by eight o'clock and cannot even sit in their own gardens after that hour. They were forbidden to visit theatres, cinema halls and any other places of entertainment. Not allowed to take part in public sports, swimming baths, tennis courts, hockey fields and other sports grounds. They cannot visit any Christians and were allowed only to go to Jewish schools, many more such restrictions. A brief account of these imposed restrictions is clearly mentioned in the initial pages of the diary:

"The rest of our family, however, felt the full impact of Hitler's anti-Jewish laws, so life was filled with anxiety. After 1940 good times rapidly fled: first the war, then the

capitulation, followed by the arrival of the Germans, which is when the sufferings of us Jews really began. Anti-Jewish decrees followed each other in quick succession. Jews must wear a yellow star, Jews must hand in their bicycles, Jews are banned from trams and are forbidden to drive, Jews are only allowed to do their shopping between three and five o'clock and then only in shops which bear the placard "Jewish shop". (pg. 20, 21)

Amidst her class mates and friends she was all alone. She never had a real friend and was always in quarrel with her mother. She doesn't want to fit into the usual slot. Wished to have her own space and always voiced her own opinions. Father was her favourite and she used to say "I can understand my friends better than my own mother – too bad! (pg 53). She always longed for someone to be her best companion and to someone to whom she can express herself. And it is until when she received a diary as a birthday gift from father, she began to express all her emotions to her best friend "Kitty" – the diary. In her own words what the diary meant for her:

"In order to enhance in my mind's eye the picture of the friend for whom I have waited so long, I don't want to set down a series of bald facts in a diary like most people do, but I want this diary itself to be my friend, and I shall call my friend Kitty. No one will grasp what I am talking about if I begin my letters to Kitty just out of the blue, so, albeit unwillingly, I will start by sketching in brief the story of my life". (pg 20)

On July 5th 1942, Anne's elder sister Margot received an official summons to report to a Nazi work camp in Germany. On July 6th they went into hiding. They were later joined by Hermann Van Pels, Otto's business partner including his wife Auguste and their teenage son Peter. They hid in the sealed off upper rooms of the annex of Otto's company building in Amsterdam. The rooms they hid in were concealed behind a moveable book case, not easily noticeable. Mrs Van Pel's dentist Fritz Pfeffe, joined them four months later. They remained hidden there for two years and one month. Anne rightly called it as their "Secret Annex". They heard about the cruelties in camps and choose to be on exile than being caught. It is right what Otto Frank said before going into hiding, that "we don't want our belongings to be seized in by the Germans, but we certainly don't want to fall into their clutches ourselves. So we shall disappear of our own accord and not wait until they come and fetch us" (pg no. 31). All kinds of thoughts disturbed her as into where they are going to hide, "in a town or the country, in a house or a cottage, when, how, and where...?" (pg 33). It was quite natural to think of all such unique possibilities as all of a sudden when one is forced to go on exile. We cannot even

imagine of such a dreadful thing, all of a sudden to leave our belongings and to go somewhere we are not sure of, whether to live or to die. All of them wore two or three layers of dress and packed just enough to hold in a sachet and left their house with anxiety. On their way they received sympathetic looks from people and their face showed how sorry they were as they couldn't help because the gaudy yellow star spoke more than needed.

“Our many Jewish friends are being taken away by the dozen. These people are treated by the Gestapo without a shred of decency, being loaded into cattle trucks and sent to Westerbork, the big Jewish camp the big Jewish camp in Drente. Westerbork sounds terrible: only one washing cubicle for a hundred people and not nearly enough lavatories. There is no separate accommodation. Men, women and children all sleep together. One hears of frightful immorality because of this; and a lot of the women, and even girls, who stay there any length of time are expecting babies.” (pg 63)

It is impossible for them to escape, most of the people in the camp are branded as inmates by their shaven heads and many also by their Jewish appearance. If it is as bad as this in Holland whatever will it be like in the distant and barbarous regions they are sent to? We can assume that most of them are murdered. The English radio speaks of their being gassed. Perhaps that is quickest way to die. We feel helpless and sympathetic for them. Anne wrote an incident which really wets our eyes: “Just recently for instance, a poor old crippled Jewess was sitting on her doorstep; she had been told to wait there by the Gestapo, who had gone to fetch a car to take her away. The poor old thing was terrified by the guns that were shooting at English planes overhead, and by the glaring beams of the searchlights. No one would dare to take her in and to undergo such a risk.” (pg 64). The Germans strike without the slightest mercy. Prominent citizens and innocent people are thrown into prison to await their fate. If the saboteur can't be traced, the Gestapo simply put about five hostages against the wall. Announcements of their deaths appear in the papers frequently. These outrages are described as “fatal accidents” and countless people have gone to a terrible fate. Evening after evening the green and grey army lorries trundle past. The Germans ring at every front door to inquire if there are any Jews living in the house. No one has a chance of evading them unless one goes into hiding. Often, they go around with lists, and only ring when they know they can get a good haul. No one is spared not even the old people, babies, expectant mothers, the sick all join in the march towards death. Their nationality and even their very existence is being questioned. The fault is them is that they were born as Jews.

“Nice people, the Germans! To think that I was once one of them too! No, Hitler took away our nationality long ago. In fact, Germans and Jews are the greatest enemies in the world” (pg 65)

It is only on the second day of arrival Anne started writing her diary. It was about how she felt on hiding and the peculiar place and its ambience. “Then I had a chance, for the first time since our arrival, to tell you all about it, and at the same time to realize myself what had actually happened to me and what was still going to happen” (pg 40). There was no variety in thoughts or nothing new to be done. They go round and round like a roundabout – from Jews to food and from food to politics. It isn’t an easy task to go on hiding with such alerts outside. Day and night more of those poor miserable people are being dragged off, with nothing but a rucksack and a little money. On the way they are deprived even of these possessions. Families are torn apart, the men, women and children all being separated. Children coming from school find that their parents have disappeared. Women return from shopping to find their homes shut up and their families gone. Every night hundreds of planes fly over Holland and go to German towns, where the earth is ploughed up by their bombs, and every hour hundreds and thousands of people are killed in Russia and Africa. No one is able to keep out of it, the whole globe is waging war and although it is going better for the Allies, the end is not yet in sight. And as for us, we are fortunate. Yes, we are luckier than millions of people. The children here run about in just a thin blouse and clogs, no coat, no hat, no stockings and no one helps them. Their tummies are empty, they chew an old carrot to stay the pangs, go from their cold homes out into the cold street and when they get to school, find themselves in an even colder classroom. Countless children stop the passers-by and beg for a piece of bread. There is nothing we can do but wait as calmly as we can till the misery comes to an end. Jews and Christians wit, the whole earth waits, and there are many who wait for death. It was not easy to go on hiding because of the terrific things happened outside, an account of what was happening outside:

“We had a short circuit last evening, and on top of that the guns kept banging away all the time. I still haven’t got over my fear of everything connected with shooting and planes, and I creep into Daddy’s bed nearly every night for comfort. I know it is very childish but you don’t know what it is like. The A.A. guns roar so loudly that you can’t hear yourself speak. Mrs. Van Daan, the fatalist, was nearly crying, and said in a very timid little voice, “Oh, it is so unpleasant! Oh, they are shooting so hard,” by which she really means am so frightened.” (pg 100)

Apart from that there were frequent banging outside that feared everyone and Anne gathered all her belongings together. She packed a suitcase with the most necessary things for an escape. But as her mother rightly said “Where will you escape to?” (pg114). They are Jews, can't go anywhere. Even the nature, birds, animals are free to do as they like but not them. They are even degraded to that level. On exile they have only option to divert their mind books, reading and studying new things. “Ordinary people simply don't know what books mean to us, shut up here, reading, learning and the radio are our amusements.” (pg 121). With the little ration they receive from fake cards they moved on. Celebrated birthday's with whatever they find. One such poem written by Margot on Anne's birthday tells us how their daily life and thoughts have been immersed in fear and agonies.

“The first shot sounds at dead of night

Hush, look! A door creaks open wide,

A little girl glides into sight,

Clasping a pillow to her side” (pg 136)

Not just that Anne used to swallow Valerian pills every day against worry and depression, but that doesn't prevent her from being even more miserable the next day. She wrote “a good hearty laugh would help more than ten Valerian pills, but we have almost forgotten how to laugh. I feel afraid sometimes that from having to be so serious I will grow a long face and my mouth will droop at the corners”. (pg 150). She wrote about the ambience there to be so oppressive and sleepy and as heavy as lead. They can't hear a single bird singing outside and a deadly close silence hangs everywhere, catching hold of them as if it will drag them down deep into an underworld. She used to wander from one room to another, downstairs and up again, feeling like a song bird whose wings have been clipped and who is hurling himself in utter darkness against the bars of his cage. She longed to “Go outside, laugh, and take a breath of fresh air, a voice cries within me, but I don't even feel a response anymore; I go and lie on divan and sleep, to make the time pass more quickly and the stillness and the terrible fear, because there is no way of killing them” (pg 155). We could understand what she needs:

“When someone comes in from outside, with the wind in their clothes and the cold on their faces, then I could bury my head in the blankets to stop myself thinking: “When will we be granted the privilege of smelling fresh air?” And because I must not bury my head in the

blankets, but the thoughts will come. Believe me, if you have been shut up for a year and a half, it can get too much for you some days. In spite of all the justice and thankfulness you can't crush your feelings. Crying, dancing, whistling, looking out into the world, feeling young, to know that am free – that is what I long for, still I must not show it, because I sometimes think is all eight of a us began to pity ourselves or went about with discontented faces where would it lead us? I couldn't talk this to anyone but only van cry. Crying can bring such relief” (pg 168)

There are a number of organisations such as “The Free Netherlanda” which forge identity cards, supply money to people “underground”, find hiding places for people, and work for young men in hiding and it is amazing how much noble, unselfish work these people are doing, risking their own lives to help and save others. Our helpers are a very good example. They have pulled us through up till-now and we hope they will bring us safely to dry land. Otherwise, they will have to share the same fate as the many others who are being searched for. Never had they heard one word of the burden which they certainly must be to them, never has one of them complained of all the trouble we give. They put on the brightest possible faces, bring flowers and presents for birthdays and bank holidays are always ready to help and do all they can. That is something we must never forget; although the Germans our helpers display heroism in their cheerfulness and affection “(pg 195). Rauter, one of the German big shots, has made a speech. “All Jews must be out of the German occupied countries before July 1. Between April 1 and May 1 the province of Utrecht must be cleaned out (as if Jews are cockroaches). Between May 1 and June 1 the provinces of North and South Holland.”(pg 108). We cannot even imagine to face such a dreadful situation. On 29th march 1944, she heard a London radio broadcast made by the exiled Dutch minister for education, art and science Gerrit Bolkestein, calling for the preservation of “ordinary documents – a diary, letters....simple everyday material” to create an archive for posterity as testimony to the suffering of civilians during the Nazi occupation. That is when she began to write more seriously that someone may read it. In August 1944, they were discovered and deported to Nazi concentration camps and that is what she heard of her last. As she said she is living in many minds even after her death and too years and years apart. It is really relevant what she said, we also need to do that to keep our minds engaged during these lockdown days.

“I finally realized that I must do my school work to keep from being ignorant, to get on in life, to become a journalist, because that is what I want! I know I can write.....but it remains to be seen whether I really have talent.....”

Exploring the Narratives of Human Resilience in History and Highlighting their Significance in Present Times
as in Anne Frank's *The Diary of a Young Girl*

Works Cited:

Frank, Anne. *The Diary of a Young Girl*: Lexicon Books. 2011

Abrams, M.H. *A Glossary of Literary Terms*. Canada: Wadson Cengage Learning, 2009s



Nature and environmentalism: Post-colonial eco critical rereading of selected Nigerian poems

Divya MS

Assistant Professor, Department of English, SCMS School of Engineering and Technology, Ernakulam, Kerala, India

Abstract

This paper is an attempt to discuss about ecology and environmentalism in the selected poems of Nigerian poets Wole Soyinka, Tanure Ojaide and Niyi Osundare in a Post-colonial Eco critical review. In literature, ecocriticism is a mode of aesthetics that deals with the nature of relation between literature and the natural environment. Its adherents investigate human attitudes towards the world as reflected in writing about nature. It is a diverse genre known by many names, including green cultural studies, eco poetics and literary analysis of the environmental. The study seeks to explore selected poems in Nigerian literature from an Eco critical perspective. The relationship between man, the environment and nature is documented in literature. Eco-critical insights are studied in the poetry of Wole Soyinka, Tanure Ojaide and Niyi Osundare. Literature resides where creation exists, and where nature exists, life exists. Literature is an imperative tool for having a historical understanding of the relationship between man and also for determining the way man treats nature in future. In the 1990's, ecocriticism gained significant prominence in the Western academia as a domain of literary research. This does not, however, indicate that the literature of earlier periods ignored ecologically conscious concerns. Similarly, ecological scepticism seeks to explain how nature is expressed in literature and how the meaning of nature and the relationship between man and nature have changed over time as they are perceived in literature. In recent decades, the natural environment has progressively become threatened by man's activities. The chosen poems are full of varied environmental details. The poets responded to their plight in distinctive perspectives through their poetry. Extreme ecological issues such as global warming, increased pollution levels, recurrent coastal flooding, tsunami and cyclones, earthquakes and floods have culminated from the incessant cutting of trees for human use and deforestation, the use of weapons and arms, radioactive elements in nuclear power plants, industrial pollution and many more. Not only has this disruption to nature caused a catastrophic change in the atmospheric conditions around the world, but the ozone layer, our earth's defensive shield, has also been destructive. And now there is a growing and crucial need to conserve our environment and make our earth a better place to live. In Nigerian Literature, the study provides a more detailed introduction to the Eco theory from its beginnings to the present. It will also address the relationship between nature and culture, the gradual progression of ecocriticism, and its related concepts.

Keywords: ecocriticism, eco psychology, eco poetics, ecological issues

Introduction

Ecocriticism is literature is an analytical method that examines the importance of the relationship between literature and the natural environment. With several names, green cultural studies, eco poetics and environmental literary criticism, it is a diverse genre. Ecocriticism began to gain prominence in Western academia in the 1990s as a sphere of literary research. Ecological criticism seeks to analyse how nature is presented in literature and how, as seen in literature, both the interpretation of nature and the relationship between man and nature have grown over time. British colonial rulers formed a chain of command in many British colonies, such as Anglo-Egyptian Sudan and Nigeria, in which colonial officials ruled over indigenous African leaders, who then governed the majority of the African indigenous population. Colonialism in Africa is primarily responsible for the continent's lack of cultural, social, and political development. The so-called empirical scrutiny of agricultural practices imposed in northern Nigerian communities by successive British colonial era authorities is an example of a European influenced paradigm pursued by African elites. Irrigation, forest

management, and extensive use of chemical fertilizers were emphasized by the colonial scientific scrutiny system. The system provided very little benefit for the region from economic development and disrupted the traditional farming practices that for centuries had sustained the local population. Researchers and academic investigators have largely overlooked the effects of postcolonial Nigeria's economic growth. The colonization process resulted in the realignment of power, with European trading companies imposed by the colonial authority replacing the hitherto domestic Nigerian authority centers such as Opobo's Ja Ja, Oguta's Kalabari and Ibadan's Ijebu.

"Ecocriticism speaks for the earth by rendering an account of the indebtedness of culture to nature while acknowledging the role of language in shaping the view of the world"

(Campbell 5)

Thus Ecocriticism begins from the conviction that the arts of creativity and the research there of will make a major contribution to the understanding of environmental issues and the various types of eco-degradation affecting planet Earth today. Global warming, which triggers rapid climate change

as a result of unequal human interactions with nature, is a real concern that marked the end of the twentieth and early twenty-first centuries. Ecological issues are caused by climate change and have become an important concern for interdisciplinary / multidisciplinary studies. Under the concept of ecocriticism, multiple literature disciplines have embraced this style of work, centred on ecological issues. The ambivalent relationships between man and nature are old and either require or need to overcome and master human romantic devotion to nature. In the foreseeable future, climate change has arisen from these anthropocentric relationships. The reality of climate change is threatening every corner of the world. Yet he believes that lethal silence is a big impediment to resolving and mitigating climate change problems. Wangari Maathai is unveiling the true global warming issues that would have dramatic consequences on Africa. At the global stage, the query is answered as:

“Africa is the continent that will hit hardest by the climate change. Unpredictable rains and floods, prolonged droughts, subsequent crop failures and rapid desertification, among other signs of global warming, have in fact already begun to change the face of Africa.”

(as cited by Toulmin, 2008, p. 1).

In environmental concerns and philosophies, there are several expressions that share similar denominators in the objective of environmental conservation. For Graham Huggan and Helen Tiffan:

“Postcolonial ecocriticism and Ecocriticism are hedged about with seemingly insurmountable problems. The two fields are notoriously difficult to define not least by their own practitioners.... Thus, internal divisions...e.g. the commitment to social and environmental justice or differences... and large scale distinctions based on the attractive view that postcolonial studies and eco/environmental studies offer mutual correctives to each other turn out to... be perilous” (3). Postcolonial ecocriticism, on the other hand, is a plurality of ecocriticism that discusses: “concerns with conquest, colonisation, racism, sexism along with its investments in theories of indigeneity and diaspora and the relations between native and invader, societies and cultures” (Huggan and Tiffan 6) to explicate Eco critical modes of feminist ecocriticism, romantic ecocriticism and postcolonial Ecocriticism “need to be understood as particular ways of reading” (Huggan and Tiffan 13). Regardless of the numerous discourses on ecocriticism and postcolonial ecocriticism, this research indicates that postcolonial ecocriticism cannot be evaluated without delving into environmental problems, and ecocriticism or eco-environmental studies cannot be discussed without discussing postcolonial concerns alongside imperialism, a metaphor that examines ideologies of supremacy and socio-history.

It is in this regard that am going to analyse some ecological problems in Wole Soyinka's poem “*Dedication for Moremi 1963*” with a post-colonial perspective. The concept of the poem is about the natural order of things, and also about bringing a child into the world. It begins with the consummation of the child, and then the birth of the child into the universe of this child, a miracle created by love. It's almost like a prayer to the Earth, and a dedication to the child. It

speaks of our human life as a whole, and also of our journey back to earth. He makes use of many poetic devices in the poem, including metaphors and a lot of imagery. The line in which he says, “your tongue arch / to scorpion tail.” is one instance that stands out as a good metaphor. A pretty metaphor compares a crying baby's tongue at birth to a scorpion tail when it flicks in terror when feeling threatened. It gives us this impression of the child being born with a venomous tongue, which later brings trouble to the parents- as well as presenting this picture of a baby's squirming tongue as it clears its lungs and wails in fear of being so unexpectedly brought into this world. There are plenty of imagery examples, including the moment where he says, “Earth's honeyed milk, wine of the only rib / Now roll your tongue into honey until your cheeks are / Swarming honeycombs — your world needs sweetening kids. Through this, we get this image of taste and touch and sight all in one, the very thought makes my mouth water. The poem is full of deep inner meanings that invoke a radiant feeling, make us wonder what it means, see these peculiar literal images that attack our senses, and give us the emotions that the poet wants us to experience. The tone of this poem is joy and wonder at the birth of a child, and all those involved can feel the spiritual journey. He relates this miracle of life to the earth, as a woman bears a child, and her fruits are brought forth by it. The sound is gentle and ties us to the earth, as if every part of this birth was nature, just like every part of any animal or plant birth. In many of his words, like baobab, roots, rain, plumb her deep for life, season, fruits, and embrace, he creates the earthy and joyful sound. They all give us the feeling of a warm earth coming together to bring this happy occasion to life. In the midst of the independence of Nigeria, Soyinka recalls the many events that took place throughout his life, such as the birth of his daughter and the opening of the first National Park in Nigeria. Soyinka writes through many frames that the poem can be read through, one being a nourishing tone for his daughter, as well as one that protects the earth and its resources. The earth can be seen as a symbol of the daughter and the daughter can be seen as a symbol of the earth. Poet gives an insight to his daughter regarding the endless parallels and metaphors about the world, and how it functions. He says, “my child- your tongue arch to scorpion tail, spit straight and return to danger's threats yet coo with the brown pigeon, tendril dew between your lips.” This is the example of Soyinka asking his daughter to be as sharp and dangerous as a scorpion but also to be caring, gentle and kind as a pigeon. He clearly shows the paternal qualities he imparts to his daughter in a manner similar to the way he tells the people of Nigeria to protect their new park. He wraps up the poem with the idea that we too must let the world depend on us in the same way we rely so heavily on the sun. We have to give earth back in the way it gives us. Soyinka evokes the past not as a dead past, but as a living one whose positive or negative results catch the present and influence the future, not historical but archetypal any more. Either to condemn those suicidal attitudes or to laud the current resistant wilderness, he evokes pastoral imagery, recalls the less anthropocentric past as a less troubled model, and projects a green future as a common dream. As the only way to face fundamental and sustainable growth, Soyinka urges readers

and listeners to take on the soil. As an expression of inextricable human ties with it, this communion with one's land at every level includes mind-set, commitment, love, and respect for oneself and all of its inhabitants. As a result of technical and scientific developments, the African holistic world view that imperialists saw as "savage" has become the global solution to the danger that climate change presents today. It's not too religious to ask "who was wild and who was civilized" if the "savage" incriminated African world view has since become a "worldwide genius" response to the climate change problem.

Tanure Ojaide is a significant Literary voice of Nigerian post-war poetry, distinguished by his recourse to the orator of his birthplace. Ojaide takes oratory as a locus of an esthetic that is conscious of rural people's arts and politics, particularly in the face of a viperous, modernity-driven establishment. The focus of his poetry on orality implies its rootedness in nature. But the point that nature in Ojaide's poetry is not merely evoked as an esthetic technique, an embellishment of what many have regarded in his poetry as an overwhelming political theme, is much more crucial to this paper. Nature is also addressed as home (the natural world, biodiversity, flora and fauna), now a forgotten home in the face of modernity and global petrodollar capitalism. In the sense of postcolonial ecocriticism, I try to point out from a reading of his poetry that the nature (environment) of the Niger Delta area from which the poet comes from is a victim of exploitation and injustice caused by large-scale oil extraction in the region, just like the people living in it; and it is no longer the pristine home it used to be. Tanure Ojaide's fifteenth poetry book, "The Tale of the Harmattan" (2007), offers poetry readers and those familiar with his work a critical insight into the Niger Delta region's bleak socio-political and economic circumstances. The plurality of the poet's concerns are oil extraction and its negative environmental and human family effects. The poems differ in style and form; however, what makes the collection a publication of substance is the poet's ability to discuss contemporary problems with a spectator's eyes, and the sincerity of an empathically inspired one. This compilation illustrates the degradation of the biodiversity and climate of the Niger Delta as a result of the extraction of oil and the marginalization of the ethnic minority in whose territories the oil is mined. In one poetry collection divided into three parts with a glossary that familiarizes the reader with the landscape, politics, Urhobo mythology, and various historical and mythical figures of Nigeria, the prolific Nigerian scholar-poet Tanure Ojaide uses bold rhetoric and a variety of techniques to claim the person of the poet as an eyewitness to historical events, especially the destruction of the destruction of the Niger Delta's ecosystem and environment as a result of oil exploitation and the marginalization of the ethnic minority people in whose land oil is exploited. He shows concern for the underprivileged and oppressed in society, whose fight for equality, fairness and justice he supports, in the course of this poetic story. Conscious of the postcolonial situation in Nigeria, his native nation, he condemns the rampant corruption that drains the country's enormous wealth. Affirming humanity, he condemns the perpetrators of genocide, as in the Darfur region of Sudan, in the strongest

possible words. The fact that what happens in Nigeria's troubled oil-rich yet poor Niger Delta region affects the worldwide price of oil demonstrates the degree of local and global connectivity, what is now described as 'glocal.' The Harmattan Tale (2007) argues that his research on the indigenous peoples (especially women) of Nigeria's Niger Delta offers an important way to revise our understanding of postcolonial theory in order to step beyond the outdated notion of colonial nations to colonialist power as sitting in multinational corporations that transcend national origin. My research combines elements from environmental, political, and socio-cultural images to analyze how Ojaide's work exposes the relationship between environmental problems and government collusion with multinational corporations, while calling for a vision of environmental justice to be accomplished by the movement of the Delta people. Ojaide's definition of historic environmental destruction and devastating oil contamination caused by multinational oil firms in the Niger Delta region is part of an interdisciplinary and multi-theoretical view of neo-colonial literature. The dialogic development of a variety of discourses is part of his complex literary style; his work involves feminist discourse and eco-critical interpretation of environmental issues, as well as post-colonial discourse that has become a defining feature of contemporary African literature. Ojaide's earlier-generation poetry and establishes him in post-colonial African poetry as a significant voice. The poems in *The Harmattan Tale* share Ojaide's love for exploring ancient African folklore with readers. In these poems, Ojaide's concerns owe much of their connection to his sensibilities and affinities towards his homeland. He does not surrender his creative inclinations or call for a Marxist agenda for political sloganeering or writing poetry, as one can admit, unaware of the genius of his imaginary complexity.

The fourth collection of poetry "The Eye of the Earth" by Niyi Ariyoosu Osundare (1986) ^[10], Nigerian ecology is celebrated in this work and focus is given to the common man where it portrays one of the fiercest indictments of the people and alien destructive powers of modern economic culture. *The Eye of the Earth* (1986) by Osundare is divided into three sections: back to earth, eye-ful glances of rain songs and home call with eighteen poems. This study investigates ecological implications in such poems as "forest echoes", "The Rocks Rose to meet me", "harvest call", "Let the earth's pain Be Soothed", "First rain", "Rain-coming", "Rain drum", "farmer-born", "They too Are the Earth", "Ours to Plough, Not to Plunder" and "Our Earth Will Not Die". *The Eye of the Earth* poetry is divided into poems of varying lengths that lament the harm to the Nigerian climate for economic reasons and technological development. The poet's memories and impressions are captured by a series of confessional and lyrical poetry. The environmental views of Osundare are drawn precisely from the Yoruba world view of traditional values taken from African culture. He claims that nature promotes a coherent equilibrium between microscopic species, insects, plants and humans and calls for the protection of the environment in Nigeria from the destruction of modern civilizations. It takes a pictorial account of man-and-earth violence. In other words, in the quest for better leadership by

alternative order, *Eye of the Earth* (1986) is dedicated to reclaiming the earth that has been forced to prostrate by capitalist processes. The poetry of Osundare is based on a vigorous, sustained concern for one of the oldest producers in the world: the peasants, those who till the land, and their quasi-mythical links to the earth. His goal is to immerse the realities and multiple lineaments of Africa's underdevelopment and poet laments on the ecological collapse and future which threatens the Nigerian landscape showing the increasing level of environmental degradation by the world's mining industries. The poet's concern for the pathetic condition of the Nigerian environment and the propensity of the Nigerian ruling class to safeguard and exploit land, power and income resources at the cost of ecological balance and the well-being of the oppressed people is self-evident in this volume of poetry. The poet is concerned with both fact and the relationship between the individual and his environment. Therefore, it is not surprising that the whole volume is dedicated to poems about man engaging with nature's physical aspects. Really, the opening poem 'Forest Echoes' is a harbinger of what's to come. The poet saunters into the Ubo Abusoro forest in the poem, from where he allows his sea of memory to flood unimpeded. The first thing that strikes the poet when he enters the forest is the destruction by timber traders of the land and the trees referred to as *agbegilodo* in the poem. From this position, the poet laments the fact that, as a consequence of exploitation, these economic trees were reduced to mere stumps. There is the palm-wine tree which is described as conqueror of rainless seasons/mother of nuts and kernels/bearer of wine and life. In 'Forest Echoes,' Osundare portrays man, the ground, animals, plants (actually all of nature) interacting and celebrating at this period of universal productivity in one festive mood. It's set in the past but it's meant to reinforce our current understanding. The second poem in the collection '*The Rocks Rose to Meet Me*' is an encounter with the rocks – another aspect of physical nature. Before the rock of Olosunta, the poet is standing and waiting like Christopher Okigbo at heavens gate. And the Olosunta rock began to address the poet in the following words:

“You have been long, very long, and far
Unwearying wayfarer,
Your feet wear the mud of distant waters
Your hems gather the bur
Of farthest forests;
I can see the west most sun
In the mirror of your wandering eyes”
(Osundare the Eye of the Earth, 13).

In these lines, Osundare is doing some kind of homecoming. He is a renegade and is now trying to establish vital links with the past. As he put it:

‘The Rocks Rose to Meet Me’ is a homecoming of a Kind, a journey back (and forth) into a receding past Which still has a right to live. The rocks celebrated in This section... occupy a central place in the cosmic Consciousness of Ikere people; they are worshipped and frequently appeased with rare gifts, thunderous

Drumming and dancing
(Osundare the Eye of The Earth ‘Preface’ xiii).

The truth is that Osundare honors the rocks of Olosunta in Ikere cosmology, since they are both an aspect of physical existence and have a supernatural dimension. It is mother earth and natural laws require that the resources of nature should be used to advance society. Osundare also revolves around the cosmology of Ikere individuals in 'Harvest Call'. The rocks that rose in the previous poem to meet the poet are also named guardians of the spirit of harvest in Ikere's worldview. Thus, in this portion of the collection, all the poems speak of crops, harvest and bounty. The assumption is that the earth is a source of development and growth. Fertile and generous, it is. It will create food and resources for the good of mankind. In fact, the earth means abundance and abundance. The Earth is seen as the centre of wealth and life. Yet the rain acts as an agent or regulator between man and Earth. In his poetry, Osundare explores and praises these two facets of nature through introspection and nostalgia. Osundare also makes the suggestion in his celebration of the theme of nature that the dispossession of the world by some powers in society is capable and can actually threaten the full life of man as a human being.

References

1. Abdu, Saleh. *The Peoples Republic: Reading the Poetry of Niyi Osundare*. Kano: Benchmark Publishers, 2003.
2. Abrams, M.H. *A Glossary of Literary Terms*. Canada: Wadson Cengage Learning, 2009.
3. Ascroft B, Gareth G, Helen T. *Postcolonial Studies: The key Concepts*. New York: Routledge, 2007.
4. Byron, Lord George ‘Childe Harold’s Pilgrimage’ Ed. Frank Kermonde and John Hollander. *The Oxford Anthology of English Literature*. Vol. 2. London: Oxford University Press, 1973, 2.
5. Barret, Lindsay. “The Niger Delta Conundrum” *New African* 483, 2009. Print. Betty Roszak and Theodore Roszak, „Deep Form in Art and Nature”. *The Green Studies Reader: From Romanticism to Ecocriticism*. Laurence Coupe (ed) New York: Routledge, 2000.
6. Bodunde, Charles. “Niyi Osundare and the Materialist Vision: A Study of the Eye of the Earth.” *Ufahamu Journal of the African Activist*, 1997; 5:81.
7. Charles E. “The Possibilities of Hope: Africa in Niyi Osundare’s Poetry”. *Lagos Papers in English* 2, 2007, 62-63
8. Chiwenzu. *Towards the Decolonisation of African Literature Vol.1*. Enugu: Fourth Dimension Publishers, 1980.
9. Edward. Said. *Culture and Imperialism*. London: Chatto and Windus, 1993.
10. Osundare, Niyi. *The Eye of the Earth*. Ibadan: Heinemann, 1986.
11. Ruecket, William. “Literature and Ecology: An Experiment in Ecocriticism” *The Ecocriticism Reader: Landmarks in Literary Ecology*. Glotfelty Cheryl and Harold Fromm. Athens: University of Georgia P, 1996.
12. Russell S. Sanders. “Speaking a Word for Nature”. *The*

- Ecocriticism Reader: Landmarks in Literary Ecology. Cheryll Glotfelty and Harold Fromm (ed). Athens: University of Georgia P, 1986.
13. Walunywa, Joseph. Postcolonial African Theory and Practice: Wole Soyinka. PhD Dissertation. Syracuse: Syracuse University, 1997.

PAPER • OPEN ACCESS

Soft morphological filtering using hypergraphs

To cite this article: Nuja M Unnikrishnan *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1085** 012038

View the [article online](#) for updates and enhancements.

You may also like

- [Dynamical systems on hypergraphs](#)
Timoteo Carletti, Duccio Fanelli and Sara Nicoletti
- [Planted hitting set recovery in hypergraphs](#)
Ilya Amburg, Jon Kleinberg and Austin R Benson
- [Analysis of quantum error correction with symmetric hypergraph states](#)
T Wagner, H Kampermann and D Bruß

A promotional banner for 'Free the Science Week 2023' featuring a dark blue background with a futuristic, glowing interface. A hand is shown interacting with a circular element containing a padlock icon. The text 'Free the Science Week 2023 April 2-9' is in the top left, 'Accelerating discovery through open access!' is in the middle left, and the ECS logo, website 'www.ecsdl.org', and a 'Discover more!' button are at the bottom left.

Free the Science Week 2023 April 2-9

Accelerating discovery through
open access!

 www.ecsdl.org [Discover more!](#)

Soft morphological filtering using hypergraphs

Nuja M Unnikrishnan¹, Mini Tom², V Bino Sebastian³ and K V Thomas⁴

¹Research Scholar, Mathematics Department, Bharata Mata College, Thrikkakara, Kerala, India, nuja.mu@gmail.com

²Professor, Basic Science and Humanities Department, SCMS School of Engineering and Technology, Karukutty, Kerala, India, minitom@scmsgroup.org

³Assistant Professor, Mathematics Department, Mar Athanasius College, Kothamangalam, Kerala, India, binosebastianv@gmail.com

⁴Associate Professor, Mathematics Department, Nirmala College, Muvattupuzha, Kerala, India, kvtbmc@gmail.com

Abstract. A new framework of soft mathematical morphology on hypergraph spaces is studied. Application in image processing for filtering objects defined in hypergraph spaces are illustrated using several soft morphological operators- openings, closings, granulometries and ASF acting (a) on the subset of vertex and hyperedge set of a hypergraph and (b) on the subhypergraphs of a hypergraph. Experimental results dealing with the extension of soft morphological operators to gray scale images are presented in this paper. The results obtained are promising and is a better substitute for the prevailing methods.

1. Introduction and Related Works

Mathematical morphology is an emerging area of image processing. In the area where image processing can be applied, mathematical morphology have provided solutions to different tasks. Soft mathematical morphology is another approach to mathematical morphology that was introduced by Koskinen et al. [1]. In this method, weighted order statistics are used instead of minima and maxima [3]. Soft morphological operators show better performance than primitive morphological operators as they are less sensitive to additive noise and to small variations in object shape [4]. Soft mathematical morphology have also been extended to fuzzy sets [3].

Morphological operators can be defined on graphs and hypergraphs. In this paper a new context is introduced that lengthens the perceptions of soft morphological filters into hypergraphs. This work is an extension of our preliminary work [2].

2. Preliminaries

In this section, we see some important definitions and properties that will be needed in the sequel.

2.1. Soft mathematical morphology using hypergraphs [2]

Hypergraph [5], [6] is defined as a pair $H = (H^*, H^X)$, where H^* is a set of points called vertices and H^X is a family of subsets of H^* called hyperedges ie; $H^X = (e_i), i \in I$ where I is a finite set of indices. The sets H^* ,



H^X and H^\bullet are the subsets of H , subsets of H^X and subhypergraphs of H respectively. Here H^X and H^\bullet are Boolean lattices. The set H of all sub hypergraphs of H forms a complete lattice [5], [6]. Morphological operators are defined on these lattices. In soft mathematical morphology, the structural element is divided into two subsets- the core and the soft boundary. In the development of the final output, weightage of core is more than the soft boundary. Hence we subdivide the hyperedges (along with the vertices belonging to it) into core B_1 and soft boundary B_2 .

2.1.1. Definition 1 [2]

The operators $\delta^\bullet, \in^\bullet$ are defined from H^X into H^\bullet and the operators δ^X, \in^X are defined from H^\bullet into H^X as follows: For any $X^\bullet \subseteq H^\bullet$ and $X^X \subseteq H^X$, where $X^X = e_i \cup e_i^*$; $e_i \in B_1$ and $e_i^* \in B_2$, $i \in J$ such that $J \subseteq I$,

$$\delta^\bullet: H^X \rightarrow H^\bullet \text{ is defined as } \delta^\bullet(X^X) = k^{th} \text{ largest of } [\cup_{j \in J} \{k \diamond v(e_j), v(e_j^*)\}; e_j \in B_1, e_j^* \in B_2]$$

$$\in^X: H^\bullet \rightarrow H^X \text{ is defined as } \in^X(X^\bullet) = k^{th} \text{ smallest of } \{k \diamond e_i, e_i^*, i \in I \mid v(e_i), v(e_i^*) \subseteq X^\bullet, e_i \in B_1, e_i^* \in B_2\}$$

$$\in^\bullet: H^X \rightarrow H^\bullet \text{ is defined as } \in^\bullet(X^X) = k^{th} \text{ smallest of } [\cap_{j \notin J} \{k \diamond \overline{v(e_j)}, \overline{v(e_j^*)}\}; e_j \in B_1, e_j^* \in B_2]$$

$$\delta^X: H^\bullet \rightarrow H^X \text{ is defined as } \delta^X(X^\bullet) = k^{th} \text{ largest of } \{k \diamond e_i, e_i^*, i \in I \mid v(e_i) \cap X^\bullet \neq \emptyset \text{ and } v(e_i^*) \cap X^\bullet \neq \emptyset; e_i \in B_1, e_i^* \in B_2\}.$$

Here, $k \diamond x$ is read as x is repeated k times. Vertex soft dilation (δ) and vertex soft erosion (\in) that act on H^\bullet , also hyper edge soft erosion (ε), hyper edge soft dilation (Δ) that act on H^X are defined as follows: The operators δ and \in that act on H^\bullet are defined by $\delta = \delta^\bullet \circ \delta^X$ and $\in = \in^\bullet \circ \in^X$ and the operators Δ and ε that act on H^X are defined by $\Delta = \delta^X \circ \delta^\bullet$ and $\varepsilon = \in^X \circ \in^\bullet$. Furthermore the operators $[\delta, \Delta]$ and $[\varepsilon, \in]$ are defined as

$$[\delta, \Delta](X) = (\delta(X^\bullet), \Delta(X^X)) \text{ and } [\varepsilon, \in](X) = (\varepsilon(X^X), \in(X^\bullet)) \text{ for any } X \in H.$$

These operators are called the soft dilation and soft erosion acting on the lattice (H, \subseteq)

3. Property

- Operators \in^X and δ^X are dual of each other. Similar duality concept hold for \in^\bullet and δ^\bullet .
- (\in^X, δ^\bullet) and (\in^\bullet, δ^X) are adjunctions.
- δ^\bullet and δ^X are soft dilations.
- \in^\bullet and \in^X are soft erosions.

Proof (a): This property is proved in [2].

(b): Let $X^X \subseteq \in^X(Y^\bullet)$. Then,

$$\begin{aligned} x \in \delta^\bullet(X^X) &\Rightarrow x \in k^{th} \text{ largest of } [\cup_{j \in J} \{k \diamond v(e_j), v(e_j^*)\}] \\ &\Rightarrow x \in v(e_j), x \in v(e_j^*) \text{ for some } j \in J \\ &\Rightarrow \exists e \in X^X \text{ such that } x \in v(e) \\ &\Rightarrow e \in \in^X(Y^\bullet) \quad \{\text{Since } X^X \subseteq \in^X(Y^\bullet)\} \\ &\Rightarrow e \in k^{th} \text{ smallest of } \{k \diamond e_i, e_i^*, i \in I \mid v(e_i), v(e_i^*) \subseteq Y^\bullet\} \\ &\Rightarrow v(e) \subseteq Y^\bullet \\ &\Rightarrow x \in Y^\bullet \quad \{\text{Since } x \in v(e)\} \end{aligned}$$

Thus, $\delta^\bullet(X^X) \subseteq Y^\bullet$.

Conversely, if $\delta^\bullet(X^X) \subseteq Y^\bullet$.

$$\begin{aligned} \text{Then, } e \in X^X &\Rightarrow v(e) \subseteq \delta^*(X^X) \\ &\Rightarrow v(e) \subseteq Y^* && \{\text{Since } \delta^*(X^X) \subseteq Y^*\} \\ &\Rightarrow e \in \in^X(Y^*) \end{aligned}$$

Therefore, $X^X \subseteq \in^X(Y^*)$.

Hence, (\in^X, δ^*) is an adjunction.

In a similar manner, (\in^*, δ^X) is an adjunction can be proved using the information that operators \in^X and δ^X are dual of each other, also \in^* and δ^* are dual of each other and (\in^X, δ^*) is an adjunction. Properties (c) and (d) follows directly using property (b).

4. Soft morphological filters

A morphological filter [5], [8] is an operator α acting on a lattice \mathcal{L} , which is increasing and idempotent. If (α, β) is an adjunction [5], [8] then α is an erosion, β is a dilation. Also, $\beta \circ \alpha$ is called an opening and $\alpha \circ \beta$ is called a closing on \mathcal{L} . Opening and closing are two commonly used filters.

4.1. Definition 2

- Soft opening $\gamma_1 = \delta \circ \varepsilon$ and closing $\Phi_1 = \varepsilon \circ \delta$.
- Soft half opening $\gamma_{1/2} = \delta^* \circ \in^X$ and half closing $\Phi_{1/2} = \in^* \circ \delta^X$.

5. Soft flat morphological Operators on weighted hypergraphs

In real phase, gray-scale soft morphological operations are difficult to work with. Threshold decomposition and stacking principle can be successfully applied on gray-scale soft morphological operations. This property allows the gray scale signals to be decomposed into binary signals and the results so obtained by processing are combined to obtain the desired gray-scale output [7]. By threshold decomposition, [8] the lattice of all subhypergraphs of H induces a lattice $\text{Fun}(H^*) \otimes \text{Fun}(H^X)$ of pairs of functions weighting respectively the vertices and the hyperedges of H such that the simultaneous threshold of these two functions at any given level produces a subhypergraph of H and the properties for hypergraph operators on the lattices H^* , H^X , or H also hold for operators on the lattices $\text{Fun}(H^*)$, $\text{Fun}(H^X)$ and $\text{Fun}(H^*) \otimes \text{Fun}(H^X)$. Thus we can define a set of soft operators which are stack analogues to the soft operators \in^X, δ^*, \in^* and δ^X defined in [2].

5.1. Definition 3.

Let $F^* \in \text{Fun}(H^*)$ and let $F^X \in \text{Fun}(H^X)$, we define

$$\begin{aligned} \delta^*(F^X)(x) &= k^{th} \text{ largest of } \{k \diamond F^X(e_i), F^X(e_i^*) \mid e_i \in B_1, e_i^* \in B_2, B_1 \cup B_2 = H^X\} \forall x \in H^* \\ \in^X(F^*)(e) &= k^{th} \text{ smallest of } \{k \diamond F^*(x), F^*(x) \mid x \in v(e), e \in B_1 \cup B_2\} \\ \in^*(F^X)(x) &= k^{th} \text{ smallest of } \{k \diamond F^X(e_i), F^X(e_i^*) \mid e_i \in B_1, e_i^* \in B_2, B_1 \cup B_2 = H^X\} \forall x \in H^* \\ \delta^X(F^*)(e) &= k^{th} \text{ smallest of } \{k \diamond F^*(x), F^*(x) \mid x \in v(e), e \in B_1 \cup B_2\}. \end{aligned}$$

By using soft flat morphology, it is possible to work with gray scale soft dilation, erosion, opening, closing, granulometries and Alternating Sequential Filters.

6. Experimental Result

To represent the hyperedges we use 4-uniform hypergraph illustrated in Figure 1. A hyperedge together with the 4 vertices belonging to it is taken as core, remaining vertices and hyperedges are taken as soft boundary. Using this simple hypergraph structure, experimental result for $k=1$ and $k=2$ are tabulated. Dilated and eroded gray scale image results are obtained to get the soft flat morphological operators defined in the previous section. Composition of these operators generates alternating sequential filters(ASF) which

are capable of removing noise effectively from binary and gray scale images. Figure 2 represent the gray scale image taken for the experimental purpose. The noisy image obtained by adding 5% salt and pepper noise is shown in figure 3. PSNR of noised image with original image is 17.24. Figure 4 and figure 5 shows the noise removed images obtained by applying alternating sequential filters (ASF) $\gamma_1 \circ \Phi_1$ and $\gamma_{1/2} \circ \Phi_{1/2}$ for $k=1$. The resultant PSNR of the noised removed images with the original image are respectively 40.53 and 40.23. Similarly figure 6 and figure 7 shows the results obtained after applying $\gamma_1 \circ \Phi_1$ and $\gamma_{1/2} \circ \Phi_{1/2}$ for $k=2$. In this case, the resultant PSNR of the noised removed images with the original image are 39.79 and 41.82 respectively. Experiment results depicts that $\gamma_{1/2} \circ \Phi_{1/2}$ for $k = 2$ is giving better approximations than $\gamma_1 \circ \Phi_1$ and $\gamma_{1/2} \circ \Phi_{1/2}$ for $k=1$.

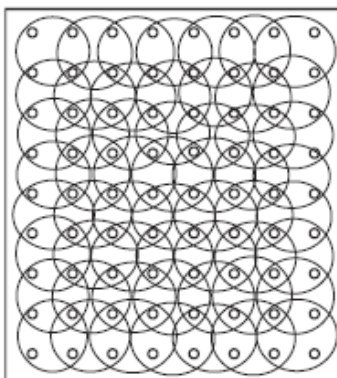

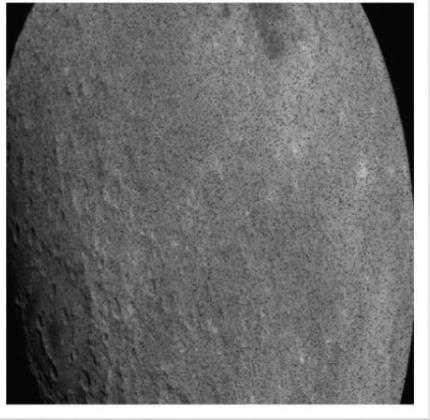






Figure 1 Four Uniform Hypergraph structure used to represent an image

PSNR of noised image with original image : 17.24	
	
Figure 2. Original Image	Figure 3. Noised Image (Added Noise 5%)

Noise Removed Images using Alternating Sequential Filters	
For K =1	
	
<p>Figure 4. $\gamma_1 \circ \Phi_1$ PSNR = 40.53</p>	<p>Figure 5. $\gamma_{1/2} \circ \Phi_{1/2}$ PSNR = 40.23</p>

Noise Removed Images using Alternating Sequential Filters	
For K = 2	
	
<p>Figure 6. $\gamma_1 \circ \Phi_1$ PSNR = 39.79</p>	<p>Figure 7. $\gamma_{1/2} \circ \Phi_{1/2}$ PSNR = 41.82</p>

7. Conclusion

The aim of this work is to recognize the prospects of using soft morphological operators under the framework of hypergraph. In our study a four uniform hypergraph structure was taken. It was then separated into core and soft boundary. But, it's a well-known fact that 3-uniform hypergraph structure gives good result for binary, gray scale and colour image filtering applications [8]. So by the choice of different hypergraph structures along with the suitable choice of core and soft boundary, results of the above method can be improved for different and bigger values of k. The initial results are promising and future work is to be done in this regard using different hypergraph structures.

References

- [1] Koskinen L, Astola J and Neuvo Y 1991 Soft morphological filters *Proc. SPIE- Int. Soc. Opt. Eng.* 262-70.
- [2] Unnikrishnan M Nuja, Sebastian V Bino and Thomas K V 2019 Soft Morphological Operators on Fuzzy Hypergraphs *Mathematical Science International Research Journal (MSIRJ)* 8 Iss.2 72-78.
- [3] Gasteratos A, Andreadis I and Ph. Tsalides 1998 Fuzzy soft mathematical morphology *IEE Proc.-Vis. Image Signal Processing.* 145 41-49.
- [4] Kuosmanen P and Astola J 1995 Soft morphological filtering *J.Math. Imag. Vis.* 5 231-62.
- [5] Sebastian V Bino, Unnikrishnan A, Balakrishnan K and Ramkumar P B 2017 Morphological filtering on hypergraphs *Discrete Applied Mathematics Elsevier* 307-20.
- [6] Sebastian V Bino, Unnikrishnan A, Balakrishnan K, and Ramkumar P B 2014 Mathematical morphology on hypergraphs using vertex-hyperedge correspondence *ISRN Discrete Mathematics.*
- [7] Shih F Y and Mitchell O K 1989 Threshold decomposition of gray-scale morphology into binary morphology *IEEE Trans. Pattern Anal. Mach. Intell.* 11 Iss.1 31-42.
- [8] Sebastian V Bino, Unnikrishnan M Nuja, Sebastian Neenu and Paul Rosebell 2020 Morphological operators on hypergraphs for colour image processing *Advanced Computing and Communication Technologies for High Performance Applications(ACCTHPA)* *IEEE Xplore* 217-20.

The hidden role Patriarchy in Malayalam Cinema: An analysis of the movie ‘sufiyum sujathayum’

Febini M Joseph¹, Cefy Joice J²

¹Assistant Professor of English, SCMS School of Engineering and Technology, Karukutty, Ernakulam, Kerala, India

²Student, Kerala Law Academy, Trivandrum, India

Received: 09 Nov 2020; Received in revised form: 23 Nov 2020; Accepted: 26 Nov 2020; Available online: 14 Dec 2020

©2020 The Author(s). Published by Infogain Publication. This is an open access article under the CC BY license

(<https://creativecommons.org/licenses/by/4.0/>).

Abstract— *Movies are the most popular medium that represents the popular taste and culture. Malayalam cinema has undergone several radical shifts throughout the past years, But still there are movies to satisfy or reinstate the traditional gender roles and patriarchy. The movie Sufiyum Sujathayum is about how society transforms our way of thinking based on the strict adherence to patriarchy. More than love people consider social acceptance as the most important priority. The paper is an analysis of the movie by using theories such as male gaze and feminism.*

Keywords— *Gender Stereotypes, Male Gaze, Patriarchy.*

Gender has been seen as a principle set up by theorists like Simone de Beauvoir. Rather than the result of sexual differences, it is represented as the consequence of social customs and practices which incorporates support from movies to practice of regular discussions as described in the book History of Sexuality by Michel Foucault. Foucault in this manner sums up that sex is the set impacts delivered on bodies, practices' and social relations by the sending of 'complex political advancements'.

Foucault's talk on the advances of sex is reformulated by Teresa De Lauretis, whose Technologies of Gender expresses that sexual orientation is a portrayal of connection having a place with a class, a gathering or a classification. The portrayal of gender starts from the principal material that contacts a child's body. The cliché garments in blue or pink breaks the perfect world and drives the youngster into a framework of portrayals and images. To cite Teresa De Lauretis, that sexual orientation isn't sex, a condition of nature however the portrayal of every person as far as a specific social connection which pre exists the individual and is predicated on the theoretical and inflexible resistance of two organic genders, which establish 'the sex-sex framework'.

Through the sexual orientation people start to fortify certain 'proper' practices, young ladies get prepared in craft and music while young men takeover the play areas. Barbie dolls and kitchen sets enhance the rooms of girls when the young boys play with automatic rifles and autos. The ongoing inclination of selecting young ladies to karate classes is furthermore just to build up their 'safeguard component' which accentuation that 'you are fragile and could be a weakling!' Our famous legends additionally strengthen the indistinguishable thought. The chivalric male warriors wandering around to abstain from squandering the moaning females, or the sovereign appeal coming to spare heaps of the alluring Rapunzel from the hands of the witch underlines the indistinguishable factor. The temperate, sensitive, delightful, and crying women are consistently princesses where on the grounds that the forceful females are consistently witches. The case of sexual orientation generalizing in legends is that the depiction of guys as globe-trotters and pioneers and females as aides or supporters.

The visual media particularly, film is one among the various innovations of sexual orientation. Laura Mulvey in her exposition, Visual Arts and Narrative Cinema clarifies the effect of visual expressions as a decent social innovation in deciding one's belief systems.

The enchantment of movies emerged from the gifted and fulfilling control of visual joys. The suggestive is being coded to the language of the prevailing male centric request in conventional standard movies and this is regularly the earlier element of achievement in each entertainment world. Movies as an assortment grasp both elitist and mainstream ideas of craftsmanship and work intimately with abstract style. The verbal and visual works of art don't appear to be only equal however intuitive and associated. A film will be considered as a social ancient rarity, which speaks to the way of life and convention to which it has a place. When it enters the social texture of a general public, it progressively impacts the way of life additionally. The entertainers likewise assume a significant job inside the methods for articulation in film.

"The visual medium offers tremendous decisions which the composed account may not. There's a more prominent opportunity inside the decision of point of view; the organizations are various camera eye, storyteller, lights, utilization of room, the language, visual correspondence, face comparatively in light of the fact that the hushes. There's likewise the vital projection of generalizations." The sex generalizations are made by these verbal and visual media's which assembles the social ideas and philosophies of the moving toward ages too. The idea of perfect spouse, perfect mother and so on are remoulded in movies. The perfect ladies in Kerala are thought as "Malayali Manka". She is considered as a kind of a goddess figure and she or he complies with each and every standards in society. She is considered on the grounds that the encapsulation of gentility. In her we will see the blooming of female temperance.

The high social improvements in Kerala lists has offered ascend to the 'fantasy of Malayali ladies 'as getting a charge out of a superior status than their partners somewhere else inside the nation, particularly the high female education inside the state. This legend has been enlarged and supported by proof that matrilineal kinds of connection designs were common in specific networks in Kerala. The elevated level of female proficiency and work, 33% reservation of seats in nearby administration bodies, high sex proportion and low fruitfulness rates alongside high female physical wellbeing accomplished a specific measure of social and political strengthening inside the property right.

Even feminine images in visual media are intended to fulfil the male looks. Intentionally or accidentally ladies emulates the vivid screen to satisfy the other gender. Malayalam film neglects to speak to the encounters of ladies from alternate point of view. At the point when a female situated film is delivered the star

esteem is a low in light of the fact that a lady assumes the principle job. Malayalam film reflects Malayali tastes, wants and dreams; one would then be compelled to surrender that so as to comprehend the contemporary public activity of karalla we ought to likewise view the delicate pornography motion pictures which once made the Kerala entertainment world drifting. Similar watchers of Adoor and Chandran films likewise delighted in Shakeela motion pictures.

The current movies or the so called movies which represent nuances in the way of presentation still gives picture of woman who are always under the control of men in the family. Obeying orders and living according to the unwritten norms are the fate of so called Malayali Mankas (A term used to represent ideal woman in Kerala) represented in movies. It is considered as usual and acceptable to everyone. But knowingly or unknowingly it provides a wrong message to the audience. The symbolic representations also denote the struggle taken by a woman when she transcends her limits. Sufiyum Sujathayum is a 2020 Malayalam movie directed and written by Naranipuzha Shanavas and produced by Vijay Babu under the banner of Friday film house. Sujatha is mute daughter of Mallikarjunan and Kamala. Sujatha was a talented dancer and an energetic girl in her village. One day she meets Sufi on her bus journey who is a disciple who returns to meet his master Ustad. Soon after their meeting both of them falls in love and they decided to elope accidentally her parents caught her and married off to Rajeev who lives in Dubai. After ten years Sufi returns to the village to meet Ustad but he was no more alive. Sufi gives out a prayer call (bank) Sufi passes away during the prayer Sujatha's husband Rajeev decides to bring her back to her village to attend Sufi's burial. Rajeev pays a visit to Sufi but Sujatha was not allowed to see him according to their beliefs woman were not allowed inside. At that evening Rajeev's passport seemed missing and they searched everywhere and he got reminded of the incident that the passport may fell into Sufi's grave and. Rajeev and his father in law decided to unearth Sufi's grave with the help of their tenant. They could not find his passport in the grave at the same time Sujatha arrives there with his passport and she throws that Misbahha (A chain with Green Beads used for prayer by Muslims).As given by Sufi gave her as her Mehar and she wanted to give him back the misbhaha that his mother gave him she placed it on his grave when her husband opened it.

The heroine is dumb and her thoughts are expressed through written words and gestures. She is lovable and everyone gives her freedom until she falls in love with a man from another religion. Her father tries to

stop her, but she plans to elope with her lover. At the moment, like several other movie scenes father tries to persuade her by sentiments. She was not able to protest and that is another symbolic way that represented the tragedy of several women. She never gets a chance to unleash her thoughts through spoken form.

The system of marriage is praised and the value is reinstated in the movie. Even though she is not mentally ready to live with her husband, she leads a troublesome life for ten long years. And the husband is always keeps jealous over her past relationship and hates her lover. When Sufi died, he ardently tries to make her realize that her love is gone forever. And when they travel together in climax scene, she holds his hands with love. And in the tomb of Sufi she throws away his ornament that she kept for all these years. It's a symbolic representation of grabbing herself from an unseen bond of love and longing.

Sujatha's grandmother was a person who was more modern in thoughts and deeds. She always respected her granddaughter's ideas and thoughts. When the groom's family came to see Sujatha, she said to them,

Avalude lokam molila...aa lokam avrude onnu kanatte
(Her world is above, let them see it too)

And when she talks with her grandmother, they discuss about a plant and her grandmother told that

"Dead bodies are buried in that place and we (woman) can't enter there. But I have gone there

These simple dialogues convey the progressive thoughts from a woman who lived a traditional life. But she deviates from the one way path of tradition. The death of grandmother is a symbolic one because it is the disappearance of a ray of hope and dreams for Sujatha.

Even though Sujatha enjoys freedom on all aspects, when she confronts with her lover or family, she sacrifices her true desires for the sake of family. Her supportive father changes completely when she is in love with a man from another religion. The conventional behaviour patterns and patriarchal ideologies are hidden while her decision making power is offended.

The movie reinstates the patriarchal ideologies that are deep rooted in Kerala. The feminine and pleasing appearance of the heroine also demands obeying and sacrificing role. At the concluding part, like a typical woman in India, she starts living in accordance with her husband. In the beginning also she awakes from a dream as if something gets dragged from their body. The various elements that are introduced contribute to reinstate the tastes of Malayali audience.

REFERENCES

- [1] Butler, Judith. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge, 1999. Print.
- [2] Beauvoir, Simone de. *The Second Sex*. New York: Vintage Books 1989, c1952. Print.
- [3] Pillai, Meena T. *Women in Malayalam Cinema: Naturalising Gender Hierarchies*. New Delhi: Orient Black Swan, 2010.
- [4] Woolf, Virginia. *A Room of One's Own*. New York: Harcourt, Brace and Company, 1929.
- [5] Foucault, Michel. *The History of Sexuality*: New York :Pantheon Books, 1978.
- [6] Bretl, Daniel J., and Joanne Cantor. "The portrayal of men and women in U.S. television commercials: A recent content analysis and trends over 15 years." *Sex Roles* 18.9-10 (1988): 595-609. Print.
- [7] Yue, Ming-Bao. "Gender and Cinema: Speaking through Images of Women." *Asian Cinema*, vol. 22, no. 1, 2012, pp. 192-207. Print
- [8] Knight, Julia. "Cinema of Women." University of Illinois Press, 2017, doi:10.5406/illinois/9780252039683.003.0017.
- [9] Devasundaram, Ashvin Immanuel. "Indian Cinema Beyond Bollywood." 2018, doi:10.4324/9781351254267.
- [10] Austin, Guy. "Representing Gender." *Algerian National Cinema*, 2019, doi:10.7765/9781526141170.000009.



RESEARCH ARTICLE

Vol. 7. Issue.1. 2020 (Jan-Mar)

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

2395-2628(Print):2349-9451(online)

A STUDY ON WOMEN EDUCATION: EDUCATE AND EMPOWER

DIVYA MS

Assistant Professor, English Department, SCMS School of Engineering and Technology, Ernakulam

Email: divya718@gmail.com



Article information

Received:27/02/2020

Accepted: 15/03/2020

Published online: 21/03/2020

doi: [10.33329/ijelr.7.1.179](https://doi.org/10.33329/ijelr.7.1.179)

ABSTRACT

Education is the key to women's empowerment. When education gets denied there isn't much to happen in her life. As usual she gets married and have children at a young age, work in unpaid or low-paying jobs, and rely on economic support from their husband or family. Without education, their future and their family's future gets limited. According to the Malala Fund, there are over 130 million girls worldwide who are not in school. Has anyone thought about them or their future? There might be an unpolished gem among them. Studies have found that if every girl completes 12 years of education, child marriage would drop by 64% and health complications from early pregnancy, like early births and child deaths, would drop by 59% and 49%, respectively. Educating women and girls also improve the economic status of a nation, which will reduce the risk of war and terrorism and can be a better harbinger of future. There are still many barriers preventing girls and women to pursue and complete their education like fees, distance or lack of transportation, being forced to work and provide for their families, being forced to marry and have children, or conflict in their hometown or country. The most difficult obstacle is the mentality of the family towards a girl child. They are denied education only because she is born as a girl. The United Nations found that as girls reach secondary school, their enrolment rates decline significantly. Only 39% of the countries have equal proportions of boys and girls enrolled in secondary education. In developing countries, 35% to 85% of girls are forced to stay at home from school to take care of their younger siblings and to manage the house while their brothers are provided with higher education. To reach the competition level and to expand their professional opportunities, women need the same experiences and skills, making post-secondary education an essential part of women's empowerment. Higher education instils them with the knowledge, competence and experience that are necessary to get involved in government, business, or even in a civil society. With higher education, women and girls have better access to health information and other beneficial services which in turn will only help the family or generation to grow and develop. We need more and more sheers to be the torch bearers of future.

Key words: Empowerment, Post-secondary education, Shero, Professional opportunities

“You educate a man; You educate a man.

You educate a woman; You educate a generation”-Brigham Young

We are living in a highly competitive world where survival is the only mantra to be focused. Such a fast growing world hardly differentiates between men and women. Education is the only key factor that brought towards such a change. Even when women have access to education, there can be many other factors that can make it difficult for her to take full advantage of those opportunities. It's women who still carries the cultural burden of being the primary homemakers and caregivers. This unpaid “second shift” means that they have less time and energy to dedicate to their studies. Family responsibilities doesn't become burden for them but that is the prime factor that is delimiting them from their higher education or carrier growth. When women become the sole providers for their families, they need to fulfil it with many other factors like being victims of domestic violence, threats like financial and from profession which becomes much more difficult for them to handle. The immense pressure that she is handling becomes worthless when it is labelled under the synonym ‘mother’ or ‘wife’. Today's era has witnessed a lot of change when it comes to single girl child nuclear family. Deciding voices are only that of their parents. Thus providing wings for them to fly high and capture their aim. Their dreams can't be in fetters and who knows there might be an unpolished gem that is being soiled. Women in India have been treated with utmost respect and dignity since time immemorial. For a nation to advance, empowering women is crucial. India was blessed with visionary women who broke the fetters of gender norms in every sector. Its long back we heard about the denial of education to women. A girl child was considered curse than a boon. Never to live in a world where mothers and sisters are dragged out of their homes and raped. Wherein she becomes both the victim and accused and its said by “them”, it's all because of her “out of the box” attitude that caused her this. Is this the caring that we need to give to our sisters, rather than rending a helping hand? We can't even skim through the news heads that is going on these days. What if such an accusation comes out, it needs to face a frequent trail from the social medias and channels who will make them even worse? Is the only solution, to remain silent and bear the tragic effects or to have a prolonged grudge and to end up in depression throughout the life? It's an open ended question towards the secularist nation. Here I would like to quote the famous wording of Michelle Obama,

“When girls are educated, their countries become stronger and more prosperous”

Let the nation realize her power of golden touch. She is to succeed where ever she goes, don't curtail her from her doing what is right. Listen to her inner voice that itself will keep the nation going. We do respect our mothers and all woman is an incarnation of her, our mother Earth. Its only she who can protect, provide, love, sacrifice selflessly. Split of the moment can we see her incarnations of a loving Sita, ready to sacrifice everything for her love and family, Durga the fearest of all who battle against evil for good. For us this is our mother, both deities in one, who loves and cares us with one hand and get furious with another. Still we love her and keep no grudges. Then how can we ever think of harming her. Where has our brains and minds gone, ready to stab the same womb from where we came? Is this what we call progression? Are we the real citizens bearing torches for a future developing nation? How can we sustain by spiting on our own veins? When will all these quires be dissolved? The only solution for all this is only education, it opens the eyes to a new world of realities and hopes for a better nation. It's clearly evident that the only means to attain empowerment is through education. Change should begin from her and its only she who can bring changes. Here I would like to cite the story of Rani Padmini a legendary 13th – 14th century Queen (Rani) of the Mewar kingdom of present day India. She was the wife of King Ratan Sen, captured and imprisoned by Delhi's sultan Alauddin Khalji. Alauddin Khalji became enamoured with Padmavati's beauty and decided to siege Chittor to obtain Padmavati. Before Chittor was captured they had to face defeat against Khalji, she and her companions committed Jauhar (self-immolation) thereby defeating Khalji's aim and protecting their honour. Coupled to the *Jauhar*, the Rajput men died fighting on the battlefield. Throughout the novel *Padmini: The Spirited Queen of Chittoor*, Mridula Behari tried to explain the immense beauty and knowledge that Padmavati possessed. It's only because of her education and knowledge, she had the courage to face such a situation where her husband the Rana itself was helpless. She was bold enough to convince everyone and to equip them for a war against the evils. There are many instances in the novel where we find her indulged in reading the ancient scriptures and getting mesmerised by that. It's the education that

brought her into a new world where she understands that action speaks more than words. There is always an urge for action than mere talks, that won't lead you anywhere. Education delimits the boundaries of caste, creed, gender, finance and what not. How true it is to quote Malala Yousafzai's wordings here,

"If one man can destroy everything, why can't one girl change it?"

As Prime Minister Narendra Modi said on the launch of the expanded 'Beti Bachao Beti Padhao' (March 8, 2018): "Daughters are not a burden but the pride of the whole family. We realise the power of our daughters when we see a woman fighter pilot. The country feels proud whenever our daughters bag gold medals, or for that matter any medal, in the Olympics." This is the time only when we think oh! are they capable of bagging a gold and that too for the nation? When they come to lime light we find applauds and cheering from all over the nation. But have ever we imagined the trials or pains that they have gone through to achieve that? Same is the case with the other gender, but they are always on safe side of the scrutiny eyes of the society. It's a boy, no matter if he is alone or late to home or traveling late or alone. The world is always open to him than to her. We need more Sheros to be an inspiration and to motivate others to come out from their shells that is encircling them towards darkness. Engaged in a combat to the march for equality with our sisters and mothers, let us understand the theme of Women's Day: "An Equal world is an enabled world: realizing women's power." After the adoption of the Beijing Agenda for Action, UN has set the year 2020 as a key year for assessing international progress towards achieving gender equality and human rights for all women and girls. The Ministry of Human Resource Development (MHRD) has much triggered up their progression under the leadership of PM Modi in providing equal opportunities. It is with immense happiness we can say that due to the Swachh Bharat Mission, 14,67,679 schools now have a functioning girl's toilet, an increase of 4.17 percentage points in comparison to 2013-14. The impact of the mission has resulted in an increase in enrolment of girls by 25 percentage points in 2018-19 from 2013-14. These figures get dimmed in a society which is indulging much more in a political game for their sustenance. We living in the safe shelters haven't ever thought about our sisters who is being victimized and marginalized. Minister of Finance Nirmala Sitharaman applauded the performance of Beti Bachao Beti Padhao in her speech on the budget: "Gross enrolment ratio of girls across all levels of education is now higher than boys. At the elementary level it is 94.32 per cent as against 89.28 per cent for boys, at the secondary level it is 81.32 per cent as compared to 78 per cent and at the higher secondary level girls have achieved a level of 59.7 per cent compared to only 57.54 per cent." The MHRD has approved 5,930 Kasturba Gandhi Balika Vidyalayas, which are girls' residential schools and have an enrolment of 6,18 lakh students, to increase equality of access and opportunity for girls. The National Incentive Scheme for Girls for Secondary Education has approved an incentive sum of Rs 8,56 crore for the 28,547 beneficiaries. According to the scheme Rs 3,000 is being deposited under the age of 16 in the name of deserving unmarried girls and entitles them to withdraw it along with interest in reaching 18 years of age and passing Class X. Besides an improvement in the girl's gross enrolment rate in schools, the educational outcomes and accomplishments have also improved.

Let's go ahead with this initiative for our sisters than to get involved in the cheap political drama that is being happening in our nation. There is considerable evidence that women's education and literacy tend to reduce the mortality rates of children. It's indeed true what Malala Yousafzai pointed out:

"We realize the importance of our voice only when we are silenced"

In accordance with the celebration of India's success in improving gender equality in the education system, much greater and collective efforts are needed to achieve the Sustainable Development Goal of eliminating gender disparities in education and ensure equal access to all levels of education and vocational training for the vulnerable, including persons with disabilities, indigenous peoples and children in vulnerable situations. History of Indian women is full of pioneers, who broke gender barriers and worked hard for their rights and made advances in politics, arts, science, law, etc. Let us cite few examples of our pioneers who made us think beyond Anandibai Gopalrao Joshi, who became the first Indian female physician in the year 1887. She was also the first Indian woman who get training in Western medicine and the first woman to travel to the United States of America. Arunima Sinha, is the first female amputee to climb Mount Everest. She is also the first Indian amputee to climb the Everest. She was a national level volleyball player who in 2011 was pushed by

robbers from a running train as she defied them. After meeting this accident, one of her legs was amputated below the knee. Arati Saha became the first Indian and Asian woman to swim across English Channel in the year 1959. She also became the first female sportsperson awarded with Padma Shri in 1960. Mother Teresa founded many Missionaries of Charity, a Roman Catholic religious congregation, giving her life to social work. Indira Gandhi became the first woman Prime Minister of India and served from 1966 to 1977. Indira Gandhi renowned as the "Woman of the Millennium" in a poll which was organised by BBC in 1999. In 1971, she became the first woman to receive the Bharat Ratna award. Justice M. Fathima Beevi became the first female judge to be appointed in the Supreme Court of India in the year 1989. In her autobiography she had said about the immense suffering that she had faced to reach such a highest peak. Her father was the only person who supported her thorough out her journey. Kalpana Chawla, the first Indian woman who reached in space. As a mission specialist and a primary robotic arm operator, she went into space in 1997.

We can move to our present Sheros starting with Mithali Raj, the first woman to score a double hundred in Test Cricket against New Zealand at Wellington, 2004. She was the first to achieve this landmark in the world. Pratibha Patil became the first woman President of India and held office from July 2007 to July 2012. Kiran Bedi, joining Indian Police Service (IPS) in 1972, she became the first woman officer in India. Moreover, later in 2003, Kiran Bedi also became the first woman to be appointed as the United Nations Civil Police adviser. Anjali Gupta is the first female flying officer in the Indian Air Force to be court martialled. She used to work for the Aircraft Systems and Testing Establishment unit in Bangalore. Anjali completed her Masters of Philosophy in Sociology from the Delhi University and was first posted at Belgaum in 2001. Sania Mirza, a professional tennis player, became the first ever Indian woman to win the Women's Tennis Association (WTA) title in 2005. Later in 2015, Sania Mirza became the first Indian woman titled as rank number one in WTA's double rankings. Saina Nehwal became the first Indian women to win a medal in Badminton at 2012 Olympic Games. Later in 2015, she became the first Indian woman to secure no. 1 position in world rankings. Mary Kom, is the only woman boxer who has won a medal in each of the six World Championships. She was the only Indian woman boxer who qualified for the 2012 Olympics and became the first Indian woman boxer to win a gold medal in Asian Games in 2014. Cited just a few but more hands of achievements are behind which are yet to be recognised and appreciated. They have set a model for us to think and act beyond.

Our government has also taken much initiative for protecting the rights of education for girls. India Post or Department of Posts, the postal system of the country, offers several savings schemes with different interest rates. The **Sukanya Samriddhi Yojana**, one such savings scheme offered by India Post, is a deposit scheme for the girl child that can be opened in any of the leading banks and post offices across the country. In such schemes the child is getting the benefit for future education and for her marriage or future life. There is a platform called WE, that is an empowerment program through which woman are trained to form self-organized and self-managed savings groups, each consisting of 15-25 members. Their aim is to develop individual empowerment and increase their access to financial resources which is the prime element for eradicating poverty. All the members meet weekly to make decisions and interact in life-skills training, discuss various issues of mutual interest. They not only give a platform for awareness but also make an effort to join together and take action to improve their lives and communities. We too need to make more and more efforts than *ME TOO* to raise our voice against the inhuman oppression and injustices that is happening worldwide. Amendments to laws are must as Judiciary is the ultimate power which we believe and rely on. For a common man judiciary is the only hope or last and final resort. We all need to respect our law than fearing it. If the administrators of law are more channelized and less corrupted, we don't have to wait for justice. It's absolutely true to say the famous phrase "Justice delayed is justice denied". Let us be the harbingers of a brighter future that initiates the slogan justice and tranquility. Let the coming era witness a world devoid of corruption, discrimination, poverty, illiteracy.

"Empower yourselves with a good education, then get out there and use that education to build a country worthy of your boundless promise"

-Michelle Obama

Bibliography

“Women empowerment”. Indian Express article. <https://indianexpress.com/article/opinion/columns/beti-bachao-beti-padhao/article/6297784>.

“First Indian Women”. India Today. <https://www.indiatoday.in/education-today/gk-current-affairs/story/the-first-indian-women-312243-2016-03-08>

“One Health and Disease: Tick-Borne.” *National Park Service*, U.S. Department of the interior, <https://www.nps.gov/articles/one-health-disease-ticks-borne.htm>.

Behari, Mridula . <https://www.amazon.in/Padmini-Spirited-Chittor-Mridula-Behari-ebook/dp/B0774NH95V>

Behari, Mridula. *Padmini: The Spirited Queen of Chittor*: Penguin publishers, 2017.
