



SCMS SCHOOL OF ENGINEERING AND TECHNOLOGY, KARUKUTTY

3.3.1 Number of research papers published per teacher in the Journals notified on UGC CARE list during 2021-2022

SL NO:	Title of paper	Name of the author/s	Department of the teacher	Name of journal	Calendar Year of publication	ISSN number	Link to the recognition in UGC enlistment of the Journal /Digital Object Identifier (doi) number		
							Link to website of the Journal	Link to article / paper / abstract of the article	Is it listed in UGC Care list
1	Malware visualization and detection using DenseNets	Anandhi V	ECE	Personal and Ubiquitous Computing	July 2021	ISSN1617-4917	https://link.springer.com/article/10.1007/s00779-021-01581-w	https://doi.org/10.1007/s00779-021-01581-w	SCI
2	'Surface Modification of Tungsten Fillers for Application in Polymer Matrix Composites	Dr Jenson Joseph	Mech	Materials Today Proceedings	2021	ISSN: 2214-7853	https://www.sciencedirect.com/journal/materials-today-proceedings	https://www.sciencedirect.com/science/article/pii/S2214785321000000	Scopus
3	Tribological and Corrosion analysis of Co-20Al-GNSA composites produced through powder metallurgy process	Dr Raghav G R	Mech	IOP Conf. Series: Materials Science and Engineering	2021	ISSN: 1757-899X	https://iopscience.iop.org/journal/1757-899X	https://iopscience.iop.org/article/10.1088/1757-899X/152/1/012030	Scopus
4	Impact of Ground Nut Shell Ash on Cobalt-Chromium metal matrix composites synthesized using Powder metallurgy process	Dr Raghav G R	Mech	IOP Conf. Series: Materials Science and Engineering	2021	ISSN: 1757-899X	https://iopscience.iop.org/journal/1757-899X	https://iopscience.iop.org/article/10.1088/1757-899X/152/1/012030	Scopus
5	In-Line and Cross-Flow Response Interactions during Vortex Induced Vibration of Marine Risers	Dr Vidyachandran	Mech	Marine Engineers Review (India)	July 2021	ISSN: 0047-5955	www.imre.in	https://www.researchgate.net/publication/350000000	Scopus
6	Effect of nanofillers on the crystalline and mechanical properties of EVACO polymer nanocomposites	Dr Gibin George	Mech	Materials Today Proceedings	April 2021	ISSN: 2214-7853	https://www.sciencedirect.com/journal/materials-today-proceedings	https://www.sciencedirect.com/science/article/pii/S2214785321000000	Scopus
7	Dispersion analysis of nanofillers and its relationship to the properties of the nanocomposites	Dr Gibin George	Mech	Materials Today Proceedings	May 2021	ISSN: 2214-7853	https://www.sciencedirect.com/journal/materials-today-proceedings	https://www.sciencedirect.com/science/article/pii/S2214785321000000	Scopus
8	Scalar and Vector Controlled Infinite Level Inverter (ILI) Topology Fed Open-Ended Three-Phase Induction Motor	Dr Jayanand B	EEE	IEEE Access	July 2021	ISSN: 2169-3536	https://ieeexplore.ieee.org/	https://ieeexplore.ieee.org/document/9470000	SCI
9	Liquefaction resistance improvement of silty sands using cyclic preloading	Dr. Akhila M	CE	IOP Conference Series: Materials Science and Engineering	2021	ISSN: 1757-899X	https://iopscience.iop.org/journal/1757-899X	https://iopscience.iop.org/article/10.1088/1757-899X/152/1/012030	Scopus
10	Effect of Plasticity of Fines on Properties of Uniformly Graded Fine Sand	Dr. Akhila M	CE	Ground Characterization and Foundations, Lecture Notes in Civil	September 2021	ISSN: 2356-2557	https://www.springer.com/series/15087	https://link.springer.com/chapter/10.1007/978-981-16413-49_1	Scopus
11	Security and Privacy in Edge-centric Intelligent Internet of Vehicles: Issues and Remedies	Dr.Varun G Menon	CSE	IEEE Consumer Electronics Magazine	September 2021	ISSN: 2152-2256	https://ieeexplore.ieee.org/xpl/RecentIssue	https://ieeexplore.ieee.org/xpl/RecentIssue	SCI
12	A Roadmap of Next-Generation Wireless Technology for 6G-enabled Vehicular Networks	Dr.Varun G Menon	CSE	IEEE Internet of Things Magazine	December 2021	ISSN: 2152-2256	https://ieeexplore.ieee.org/xpl/RecentIssue	https://ieeexplore.ieee.org/document/9710000	SCI
13	Optimal Distribution of Workloads in Cloud-Fog Architecture in Intelligent Vehicular Networks	Dr.Varun G Menon	CSE	IEEE Transactions on Intelligent Transportation Systems	July 2021	ISSN: 1558-0016	https://ieeexplore.ieee.org/xpl/RecentIssue	https://ieeexplore.ieee.org/document/9400000	SCI
14	Linked Data Processing for Human-in-the-Loop in Cyber-Physical Systems	Dr.Varun G Menon	CSE	IEEE Transactions on Computational Social Systems	October 2021	ISSN: 2329-924X	https://ieeexplore.ieee.org/xpl/RecentIssue	https://ieeexplore.ieee.org/document/9390000	SCI
15	An AI-enabled lightweight data fusion and load optimization approach for Internet of Things	Dr.Varun G Menon	CSE	Future Generation Computer Systems	September 2021	ISSN: 1872-7115	https://www.sciencedirect.com/journal/future-generation-computer-systems	https://www.sciencedirect.com/science/article/pii/S1872711521000000	SCI
16	Detection and robustness evaluation of android malware classifiers	Josna Philomina	CSE	Journal of Computer Virology and Hacking Techniques	June 2021	ISSN2197-9995	https://link.springer.com/journal/11416	https://link.springer.com/article/10.1007/s11416-021-00000-0	SCI

17	IoT Based Power Analyzer for an Automated Home	Susmi Jacob	CSE	International Research Journal of Science Engineering and	2022	ISSN 2454 -3195	http://riset.com/article	http://riset.com/article	UGC CARE
18	Image Splicing Detection-Comparsion of DMVN and DMAC Networks	Litty Koshy	CSE	International Journal of Advanced Research Trends in Engineering	August 2021	ISSN 2394-3785	https://ijarct.com/	https://ijarct.com/4101vks01aug2021kristinivoti/c	UGC CARE
19	Evaluation of adversarial machine learning tools for securing AI systems	Asha S	CSE	Journal of Cluster Computing	September 2021	ISSN1573-7543	https://www.springer.com/journal/10586	https://link.springer.com/article/10.1007/s11092-021-10001-0	SCI
20	IMPACT ANALYSIS OF THE COVID19 ON THE ATMOSPHERIC AIR QUALITY AND ELECTRICITY CONSUMPTION PER DAY IN INDIA.	Dr.Sreeleka Menon	Basic Science and Humanities	Indian Journal of Applied Research	January 2021	ISSN:2249-555X	https://www.worldwidejournals.com/indian-	https://www.worldwidejournals.com/indian-	UGC CARE
21	An Intelligent heart disease prediction system based on swarm-artificial neural network	Dr.Varun G Menon	CSE	Neural Computing and Applications	May 2021	0941-0643	https://link.springer.com/journal/521	https://link.springer.com/article/10.1007/s00541-021-00551-1	SCI
22	Exploiting benefits of IRS in wireless powered NOMA networks	Dr.Varun G Menon	CSE	IEEE Transactions on Green Communications and Networking	March 2022	2473-2400	https://ieeexplore.ieee.org/xpl/RecentIssue.jsr?number=7511293	https://ieeexplore.ieee.org/document/9686954	SCI
23	Cognitive smart cities: Challenges and trending solutions	Dr.Varun G Menon	CSE	Expert Systems	1-May-2022	ISSN:1468-0394	Expert Systems - Wiley Online Library	Combinative smart cities: Challenges and trending solutions	SCI
24	Deep Reinforcement Learning for Intelligent Service Provisioning in Software-Defined Industrial Fog Networks	Dr.Varun G Menon	CSE	IEEE Internet of Things Journal	1-Jan-2022	2327-4662	https://ieeexplore.ieee.org/xpl/RecentIssue.jsr	https://ieeexplore.ieee.org/document/9686954	SCI
25	Towards Energy-Efficient and Delay-Optimized Opportunistic Routing in Underwater Acoustic Sensor Networks for IoUT Platforms: An Overview and New Suggestions	Dr.Varun G Menon	CSE	Computational Intelligence and Neuroscience	March 2022	1687-5273	https://www.hindawi.com/journals/cin/	https://www.hindawi.com/journals/cin/2022/7061617/	SCI
Total number of research papers published per teacher in the Journals notified on UGC CARE list during 2021-2022									25



Joshi

PRINCIPAL
SCMS SCHOOL OF ENGINEERING & TECHNOLOGY
VIDYANAGAR, PALLISSERY, KARUKUTTY
ERNAKULAM, KERALA-683 576

Security in Edge-Centric Intelligent Internet of Vehicles: Issues and Remedies

Publisher: IEEE

Cite This

PDF

Mainak Adhikari ; Ambigavathi Munusamy ; Abhishek Hazra ; Varun G Menon ; Vijay Anavan... All Authors

29

Cites in
Papers

805

Full
Text Views



Need Full-Text
access to IEEE Xplore for your organization?
CONTACT IEEE TO SUBSCRIBE >

Abstract

Abstract:

Propelled by the growth of automotive industry, and the ubiquity of smart sensors, intelligent transport systems such as the Internet of Vehicles (IoV) have seen significant research interest in recent times. The emerging distributed IoV networks support real-time vehicular applications through on-device computing, communication-efficient data processing, edge computing, and cloud aggregation. While enriching the user experience by minimizing the end-to-end latency through efficient energy management, IoV deployments face the fundamental challenge of security attacks. In this article, we discuss various security attack

Authors

Figures

References

Citations

More Like This

Provably Secure Authentication

Scheme for Fog Computing

Enabled Intelligent Social

of Vehicles

IEEE Transactions on Vehicular

Technology

Published: 2024

PDF

Help

Improved Multiresolution

IEEE Transactions on Vehicular

Technology

Published: 2024

Improved Multiresolution Neural

Network for Mobility-Aware

Security and Content Caching for

Internet of Vehicles

IEEE Internet of Things Journal

Published: 2023

PDF

Help

Show More

References

Citations

Keywords

Metrics

applications through on-device computing, communication-efficient data processing, edge computing, and cloud aggregation. While enriching the user experience by minimizing the end-to-end latency through efficient energy management, IoV deployments face the fundamental challenge of security attacks. In this article, we discuss various security attack modes in an edge-centric intelligent IoV framework, consisting of distributed smart vehicles, and remote processing units. We highlight various attack detection and mitigation mechanism for the proposed IoV framework, to address the security challenges. Finally, we shed light on several future research directions to ensure security of sensory data in edge-centric IoV systems.

Published in: IEEE Consumer Electronics Magazine (Volume: 11 , Issue: 6, 01 November 2022)

Page(s): 24 - 31

DOI: 10.1109/MCE.2021.3116415

Date of Publication: 30 September 2021 ? Publisher: IEEE

ISSN Information:

Print ISSN: 2162-2248

Electronic ISSN: 2162-2256

IEEE
Get Published in the
*IEEE Systems, Man, and
Cybernetics Letters*

Activate Windows
Go to Settings to activate Windows

Effect of Plasticity of Fines on Properties of Uniformly Graded Fine Sand



M. Akhila, K. Rangaswamy, N. Sankar, and M. R. Sruthy

1 Introduction

Even though researchers separate soils based on particle size as sand, silt and clay, in the field, soil always exists as a combination of all these. There are many studies concentrating on the effect of fines on the shear characteristics of sand [1–3] and liquefaction [4–7] but only a few studies have considered the other properties.

Yang and Wei [8] have analysed the change in critical state friction angle for Fujian and Toyoura sands. For clean sand without fines, the critical state friction angle tends to decrease with increasing roundness of sand particles. When those sands were tested with fines (round shape), the critical state friction angle of the mixture tends to decrease with an increase in fines content. But for fines with an angular shape, the critical state friction angle tends to increase with fines content. Phan et al. [9] have conducted one-dimensional consolidation tests on sand–silt mixtures (with low-plastic fines at a constant void ratio and constant relative density) and indicated that the behaviour of the mixtures were similar to those of loose sand. The effect of fines on void ratios was studied by Cubrinovski and Ishihara [10]. The authors reported that the void ratio initially decreases as the fines content increases from 0–20% and above 40% fines, the maximum and minimum void ratios were seen to increase steadily.

It is clear from the literature that the studies on the effect of plasticity of fines on the properties of sand are limited. Hence, the present study is focused on the effect of the amount of fines and the type of fines (or plasticity index of fines) on various properties of sand like specific gravity, limiting void ratios, grain size characteristics, angle of internal friction and compression index.

M. Akhila (✉) · M. R. Sruthy

SCMS School of Engineering and Technology, Ernakulam, Kerala, India

K. Rangaswamy · N. Sankar

NIT Calicut, Kozhikode, Kerala, India

Liquefaction resistance improvement of silty sands using cyclic preloading

Akhila M^{1,3}, Rangaswamy K² and Sankar N²

¹Department of Civil Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala, India

²Department of Civil Engineering, NIT Calicut, Kerala, India

³E-mail: akhilam@scmsgroup.org

Abstract. Liquefaction induced damages are plenty and cause various levels of destruction to civil engineering infrastructure. It is possible to prevent liquefaction-induced hazards by understanding the mechanism and adopting some improvement techniques or design the structure to resist the soil liquefaction. In the present study, the influence of cyclic preloading on the liquefaction resistance of sand-silt mixtures is analyzed by conducting undrained cyclic triaxial tests on the cylindrical samples reconstituted at medium dense conditions ($D_r = 50\%$). All samples were tested at an effective confining pressure of 100 kPa by varying the cyclic stress ratios (CSR) in the range of 0.127 to 0.178 using a sinusoidal waveform of frequency 1 Hz. The results are presented in the forms of the pore pressure build-up, axial strain variation and liquefaction resistance curves. Test results indicate that the liquefaction resistance of silty sands is increased substantially with the application of preload under drained conditions.

1. Introduction

Liquefaction induced damages are plenty and cause various levels of destruction to civil engineering infrastructure. It is possible to prevent liquefaction-induced hazards by understanding the mechanism and adopting some improvement techniques or design the structure to resist the soil liquefaction. The first possibility is to avoid the construction on liquefiable soil deposits as far as possible. However, it is mandatory to utilize the available land for the various infrastructure developments due to scarcity in the availability of land even it does not satisfy the required properties. Hence, the second option is to make the structure resistant to liquefaction by adopting deep foundations. Nevertheless, the deep pile foundations may not prevent liquefaction damages in all cases. Piles are causing to deflect in liquefaction susceptibility zones. Hence, the third option is liquefaction mitigation which involves improving the strength, density, and drainage characteristics of the soil. The selection of the most appropriate ground improvement method for a particular application could depend on many factors including the type of soil, level, and magnitude of improvement to be attained, required depth and extent of the area to be covered. This paper presents an experimental study regarding the applicability of preloading for the improvement of liquefaction resistance.

2. Literature review

Preloading of the soils occurs naturally (for eg., erosion, the flow of groundwater, etc) or artificially (purposeful preloading to improve the soil properties, demolition of structures, etc). A few researchers have analyzed the liquefaction resistance of preloaded soils. The details are given in Table 1.



Received June 19, 2021, accepted July 6, 2021, date of publication July 9, 2021, date of current version July 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3096125

Scalar and Vector Controlled Infinite Level Inverter (ILI) Topology Fed Open-Ended Three-Phase Induction Motor

A. HAREESH¹, (Member, IEEE), AND B. JAYANAND², (Member, IEEE)

¹Electrical and Electronics Engineering Department, Government Engineering College, Thrissur, Affiliated to APJ Abdul Kalam Technological University, Thrissur, Kerala 680009, India

²Electrical and Electronics Engineering Department, SCMS School of Engineering and Technology, Affiliated to APJ Abdul Kalam Technological University, Karukutty, Kerala 683576, India

Corresponding author: A. Hareesh (harielayur@gmail.com)

ABSTRACT The design and performance analysis of an open-ended three-phase induction motor, driven by an Infinite Level Inverter (ILI) with its speed control using scalar and direct vector control techniques are presented in this paper. The ILI belongs to an Active-Front-End (AFE) Reduced-Device-Count (RDC) Multi-level Inverter (MLI) topology. The fundamental structure of this inverter topology is a dc-to-dc buck converter followed by an H-bridge. This topology performs a high-quality power conversion without any shoot-through issues and reverse recovery problems. The performance of the proposed topology is validated using a resistive load. The THD of output voltage waveform obtained is 1.2%. Moreover, this topology has exhibited a high degree of dc-source voltage utilization. ILI considerably reduces the switching and conduction losses, since only one switch per phase is operated at high frequency, and other switches are operated at power frequency. The overall efficiency of the inverter is 98%. The speed control performance of the ILI topology using three-phase open-ended induction motor has been further validated through scalar and direct vector control techniques. Results obtained from simulation studies are verified experimentally.

INDEX TERMS Active-front-end, multi-level inverters, reduced-device-count, scalar and direct vector control, three-phase infinite level inverter.

AFE	Active Front End	qZSI	Quasi Z- Source Inverters
CHB	Cascaded H- Bridge	RDC	Reduced Device Count
ESR	Equivalent Series Resistance	RSA	Reduced Switch Asymmetrical
FOC	Field Oriented Control	RSM	Reduced Switch Modified
FC	Flying Capacitor	RSS	Reduced Switch Symmetric
GaN	Gallium Nitride	Si	Silicon
IGBT	Insulated Gate Bipolar Transistor	SiC	Silicon Carbide
ILI	Infinite Level Inverter	SM	Sub Modules
M	Modulation Index	SPWM	Sinusoidal Pulse Width Modulation
MIC	Module Integrated Converter	SVPWM	Space Vector Pulse Width Modulation
MLI	Multi-Level Inverter	THD	Total Harmonic Distortion
MMC	Modular Multilevel Converter	WBG	Wide Band Gap
NPC	Neutral Point Clamped	ZVS	Zero Voltage Switching
PWM	Pulse Width Modulation	C	Filter Capacitance
qCHB-FLBI	Quasi Cascaded H- Bridge Five Level Boost Inverter	D	Diode
		f_c	Carrier Frequency
		f_m	Modulating Frequency
		f_s	Fundamental Frequency
		f_{sw}	Switching Frequency
		$I_{as}^*, I_{bs}^*, I_{cs}^*$	Stator Phase Current References

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaofeng Yang¹.

i_c	Current through the Capacitor
$I_{Cripple}$	Capacitor Ripple Current
I_{Cpeak}	Peak Current through the Capacitor
i_D	Current through the Diode
I_F^*, I_T^*	Flux And Torque Producing Components of Current
$I_{Drating}$	Current Carrying Capacity of Switch
i_L	Current through the Inductor
I_{Lavg}	Average Inductor Current
I_{Lpeak}	Peak Current through the Inductor
$I_{Lripple}$	Inductor Ripple Current
I_m	Maximum Current
i_{ph}	Phase Current
I_{rr}	Reverse Recovery Current
I_s	Stator Current
I_s^*	Phasor Reference Current
i_{SW}	Current through the Switch
$I_{SWrating}$	Current Carrying Capacity of Switch
L	Inductance
P_D	Switching Power Dissipation
$Qc1, Qc2, Qc3$	High Frequency Operating Switches
$Q1, Q2, Q3, Q4$	Power Frequency Operating Switches
R_L	Load Resistance
T_e	Actual Torque
T_e^*	Reference Torque
T_{OFF}	Turn OFF Time
T_{ON}	Turn ON Time
t_{rr}	Reverse Recovery Time
T_s	Sampling Time
V_c	Voltage across the Capacitor
V_{co}	Average Output Voltage across Buck Capacitors.
V_{ce_sat}	Collector Emitter Voltage at Saturation
V_{Cmax}	Maximum Voltage across the Capacitor
$V_{Cripple}$	Capacitor Voltage Ripple
V^D	Voltage across the Diode
$V_{Dstress}$	Voltage Stress across the Diode
V_f	Forward Voltage of Diode
V_{GS}	Gate Source Voltage
V_i	Input DC- Source Voltage.
V_L	Voltage across the Inductor
V_m	Maximum Voltage
V_o	Sinusoidal Output Voltage
V_{oa}, V_{ob}, V_{oc}	Output Phase Voltages
V_{ph}	Phase Voltage
V_s	Standing Voltage across the Switch
V_{SW}	Voltage across the Switch
$V_{SWstress}$	Switching Voltage Stress
δ	Sinusoidal Varying Modulation Index
ω	Angular Frequency
ΔI_L	Linear Current Change through Inductor
ω_r	Actual Speed
ω_r^*	Reference Speed

λ^*	Reference Rotor Flux Linkage
λ_r	Flux Linkage
θ_T	Torque Angle
\mathcal{L}	Laplace transform

I. INTRODUCTION

Nowadays dc-to-ac inverters are playing a crucial role in power electronic areas such as electric drives, electric vehicles/hybrid electric vehicles, uninterruptable power supplies, HVDC power transmission, renewable energy integration, Flexible AC Transmission Systems (FACTS) and static VAR compensators. Based on the development and the nature of output voltage waveform, the inverters are broadly classified as two-level or square wave inverters, quasi-square wave inverters, two-level PWM inverters and multilevel inverters (MLI). The major problems associated with conventional or two-level inverters include the requirement of semiconductor devices of higher power ratings. To obtain the required voltage/current capacity, many devices need to be connected in series/parallel strings. As a result, these inverters generate low power quality output waveforms along with more conduction loss. Hence to overcome the aforementioned drawbacks, MLI can be chosen as a better alternative [1].

During the past three decades, MLIs [2]–[7] attracted wide attention both in the scientific as well as in industrial fields, because of its improved power conversion capabilities such as better power quality, control and efficiency as compared to all other conventional inverters. The basic concept of MLI is to generate a higher number of voltage levels with less distortion. The power conversion is performed by various dc low voltage levels along with several low rated power semiconductor switches. Each level is defined as the portion of voltage waveform where the magnitude of voltage remains constant for a defined duration which leads to the generation of a staircase voltage waveform. If the number of voltage levels increases, then the power quality of the output voltage waveform also increases. In modern industrial applications, demand for MLIs have increased due to its viable technology to implement controlled speed drive and to maintain power quality in high-power applications. The basic fundamental topologies of MLIs are classified into three categories; they are i. Cascaded H-Bridge (CHB) [2], [8], [9], ii. Neutral Point Clamped (NPC) or Diode Clamped [10] and iii. Flying capacitor (FC) [11]. They are popularly known as classical topologies. Out of these fundamental topologies, first two are required to have a single dc source, while the third one requires multiple dc-sources. But, to realize those topologies, more numbers of semiconductor devices and passive components are required. Hence the system becomes bulky and complex. One of the most desirable properties of the multilevel structure is modular structure or modularity. Among the aforementioned topologies, the CHB due to its modularity exhibits higher output voltage, power level and reliability. Practically, NPC and FC topologies are feasible up to 5-levels only, beyond this limit, their structure and

control implementation become more complex and also the device count for different ratings increases a lot. On the other hand, for higher levels, CHB is the best solution, since its devices are of the same power rating. However, the control implementation is more complex due to the requirement of coordination among too many power semiconductor devices and the increased number of dc-sources required for generating a higher number of levels [9]. Apart from the classical topologies, several novel application-oriented MLI topological developments have been reported in various literatures. They are basically derived from classical topologies or their hybrids. However, none of the new generation topological developments can be claimed to be absolutely advantageous. Most of them are designed for specific applications.

In general, the application-oriented topologies have been focussed on increasing number of output voltage levels and power quality, with reduced number of switching devices, passive component counts and cost. Some of the application-oriented approaches are discussed in [12]–[19]. Another significant development in MLIs is fault-tolerant topology. Chen *et al.* [20] has proposed a fault-tolerant topology to obtain uncompromised multilevel voltage waveform in the event of partial failure(s) in the power circuit. Such topologies maintain the output voltage waveform utilizing control signal modification along with redundancy offered by multi switching states. Another important contribution in MLI topologies was introduced by Glinka [21] presented a new single DC source-based multilevel topology, known as ‘Modular Multilevel Converter (MMC)’. The salient features of this MMC are its modularity and scalability. This topology is simpler than CHB inverter. It could conceptually meet any voltage level requirements with reduced harmonic issues, lower converter components ratings, and also with improved efficiency. The proposed topologies have caught the attention for medium/high-power energy conversion systems, industrial applications including FACTS, HVDC transmission systems, medium-voltage variable-speed drives and medium/high voltage dc-to-dc converter applications. Suman Debnath *et al.* [22] presented a review article regarding MMC. It is highlighted with a general overview on the basics of operation, control challenges, state-of-the-art control strategies and application challenges. Majority of MLIs are originated by arranging different configurations of Sub Modules (SM), which are able to connect to each other, that can create conventional and modern MLIs. Vijeh *et al.* [23] presented a general review of MLIs based on main submodules.

In the last couple of decades, the researchers are taking efforts and are being directed to reduce the device count (RDC) in MLI topologies. The reduction of semiconductor switches and passive components in MLIs can improve the efficiency and reliability of the system as well as the overall loss, cost, size and complexity. A large number of different novel RDC-MLI topologies have been reported in different literatures [24]–[34]. The aforementioned topologies have their own merits and demerits from application requirements.

Bana *et al.* [35] presented a comprehensive review on some recently developed RDC-MLI topologies, which are more suitable in different applications such as machine drives, renewable energy systems and FACTS. These topologies can be used in grid-tied as well as in standalone applications. The RDC-MLI topologies are broadly classified into three types. They are reduced switch symmetric H-bridge type MLI (RSS-MLI) [36], [37] and reduced switch asymmetrical H-bridge type MLI (RSA-MLI) [38]. The general structure of the configurations mentioned above is an RDC-MLI coupled with an H-bridge, whereas reduced switch modified MLI (RSM-MLI) [39]–[41] topologies are without H-bridge.

In recent years, a number of novel topologies have been noticed in various literatures. The general structure of such existing topologies are Active Front End (AFE)-RDC-MLI topology. These existing MLI topologies varies depending on the dc-to-dc converter used. Those topologies consist of an AFE converter (dc-to-dc converter) followed by a synchronized H-bridge. The main role of the AFE converter is generation of voltage levels, and that of H-bridge topology is polarity generation (inverter). Different AFE- RDC-MLI topologies are discussed in various literatures [42]–[53].

This paper presents the design and performance analysis of a three-phase AFE-RDC-MLI topology for open-ended three-phase induction motor drive application. In this proposed topology dc-to-dc buck converter is used for generating different voltage levels, which is operating at high switching frequency. The generation of output voltage levels depends upon the switching frequency of the AFE converter. If the switching frequency is increased, then the number of output voltage levels also increases. Due to the higher number of voltage levels, the output voltage waveform becomes smooth sinusoidal. If the switching frequency of the AFE converter is increased to almost infinity, the levels generated in the output voltage waveform becomes infinite. Hence, the power quality of the output voltage waveform is enhanced. Therefore, the proposed topology is called as an infinite level inverter (ILI) topology. The performance of the proposed topology has been validated with a resistive load as well as with an open-ended three-phase induction machine. Results obtained from the simulation studies are verified experimentally. The taxonomy of MLIs are briefly discussed in TABLE-1. The proposed MLI topology has several advantages.

- 1) Only one switch per phase operates at high frequency.
- 2) H-bridge inverter circuit operates at power frequency.
- 3) This topology is free from shoot through menace because of the presence of inductors between the voltage source and inverters.
- 4) The output filtering capacitors in AFE converters can be an electrolytic capacitor which is smaller and less expensive as compared to ac capacitors.
- 5) This topology is suitable for implementing sophisticated algorithms.
- 6) The advance control strategies developed by the researchers related to dc-to-dc converter can be directly

TABLE 1. Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
1	[1], Feb, 2020.	Review paper; Reduced Switch Count (RDC) MLI	A comprehensive review of some of the recently proposed RDC -MLI topologies. Pd: Total power semiconductor switch count, number of DC sources, passive component requirement, highest switch voltage rating, total standing voltage	A qualitative and quantitative summary of various new MLI topologies have also been discussed and a comparison is made.
2	[2], Aug, 2002	Diode-clamped, capacitor-clamped and cascaded multicell with separate dc-sources	A survey of topologies, controls, and applications. Pd: High-voltage high-power devices, optical sensors and other opportunities for future development, Control and modulation methods.	Different industrial applications and technological aspects are discussed
3	[3], Jan/Feb 1999	Six-level back-to-back 10-kW diode-clamped converter prototype	The multilevel diode-clamped converter; obtains well-balanced voltages across the dc-link capacitor with less THD. Pd: THD, efficiency and power factor.	Simulation and experiment performed using 32-bit digital signal processor. Application: high-power and/or high-voltage electric motor drives.
4	[4], May/June, 1996	Diode-clamp, flying capacitors and cascaded-inverters with separate dc-sources.	The operating principle, features, constraints, and potential applications of these converters will be discussed. Pd: Harmonics, EM1 problems, voltage stress, components and topology comparison.	Simulation and experiment. Application: Reactive power compensation, back-to-back inertia, utility compatible adjustable speed drives.
5	[5], Aug, 2010	Review paper; Multilevel converters.	Recent contributions on topologies, modulation, and recent advances and industrial applications of multilevel converters. Pd: Current state in industrial applications, Modulation and control of multilevel converters, Nontraditional applications of MLI, Future trends and challenges	High-performance multilevel converter applications based on topologies are tabulated.
6	[6], July, 2010	A survey - Cascaded MLIs	A survey of different topologies, control strategies and modulation techniques used by cascaded MLI. Pd: High degree of modularity, the possibility of connecting directly to medium voltage, high power quality, both input and output, high availability, and control of power flow in the regenerative version.	Application: Pumps, fans, STATCOM, traction, LNG plant.
7	[7], June, 2008	A Review: Potential in current and future power applications	Review and analyse the most relevant characteristics of MLI and provides an introduction of the modelling techniques, operational and technological issues. Pd: Modeling techniques, modulation strategies, operational and technological issues.	Multilevel converter-driven applications overview.
8	[8], Dec, 2007	Review; Voltage-source-converter topologies. 1) 2L VSCs, 2) NPC, 3) CHB, and 4) FC multilevel VSCs.	Review of voltage-source-converter topologies. Pd: Operating principle of each topology and Modulation methods. The latest advances and future trends of the technology are discussed.	Application: Industrial medium-voltage drives.
9	[9], Oct, 2002	Cascade multilevel converter	Charge balance control schemes used for cascade multilevel converter in order to maintain equal charge/discharge rates from the dc sources (batteries, capacitors, or fuel cells) in an HEV. Pd: Output voltage, THD , efficiency and power factor.	Simulation and experimental. Application: Hybrid electric vehicles
10	[10], July, 2000	Diode Clamping MLI	Diode clamping inverter, which works without the series association of the clamping diodes. Pd: Blocking Voltage Distribution, switch and diode clamping mechanism, auxiliary clamping, operation principle.	Experimental. Application: Diode clamping MLI in large power conversion area.

TABLE 1. (Continued.) Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
11	[11], Aug, 2002	Flying Capacitor MLIs	The requirements imposed by a direct torque control (DTC) strategy on multilevel inverters are analyzed. Pd: Flux and torque.	Simulation and experimental. Application: Motor drive
12	[12], Jan, 2021	A Novel Asymmetrical 21-Level Inverter for Solar PV Energy System With Reduced Switch Count	The proposed topology achieves 21-level output voltage without H-bridge using asymmetric DC sources. Pd: RDC, THD, maximum power point tracking, total standing voltage, switches count, and sources count, gate driver boards, the number of diodes and capacitor count and component count level factor.	Simulation and experimental. Application: Solar PV Energy System
13	[13], 2020.	Hybrid multilevel dc-to-ac converter using single-double source unit	The hybrid topologies compared with the conventional CHB converter, and the best topologies recommended for medium voltage applications. Pd: Number of power switches, THD, symmetric and asymmetric inverter, high and low voltage comparison included.	Experimental. Application: Medium voltage applications. Grid-connected photovoltaic system.
14	[14], 2020.	Transistor clamped T-type multilevel H-bridge inverter	Investigation of a transistor clamped T type H-bridge MLI with Inverted double reference single carrier PWM technique (IDRSCPWM) are discussed. Pd: RDC topology , THD	Experimental. Application: Renewable energy applications
15	[15], 2020.	Cascaded MLI	Design and implementation of MLIs for fuel cell energy conversion system. Pd: RDC topology, THD, quality of output and stress, efficiency and power loss.	Simulation and experimental. Application: Fuel cell energy conversion system
16	[16], 2020.	Binary hybrid MLI	The DSOGI controller is implemented to control a binary hybrid MLI by which real power control, reactive power control, harmonic elimination and source current balancing. Pd: RDC, THD, efficiency, Power quality.	Simulation and experimental. Application: Grid integrated solar energy conversion system.
17	[17], 2017.	Cascaded H-bridge MLI	Design and testing of 5-level symmetric cascaded MLI-CHB inverter with 6 switches for photovoltaic system. Pd: THD, power loss, implementation complexity, cost, advantages.	Simulation and experimental. Application: Photovoltaic system.
18	[18], 2020.	Three Phase MLI	A new three phase MLI with reduced number of components count is proposed. Pd: RDC, THD, compared number of levels, active switches, diodes, capacitors and total standing voltage.	Simulation and experimental. Application: Low and medium power photovoltaic systems.
19	[19], 2020.	Dual source MLI (DS-MLI)	DS-MLI with fewer power switches are proposed for solar PV power conversion systems. Pd: RDC, THD, efficiency, loss, Small signal model is discussed. The proposed one is compared with its conventional counterpart.	Simulation and experimental. Application: Domestic solar PV applications
20	[20], Mar, 2005.	Multilevel converter topology	A multilevel converter topology with fault-tolerant ability is presented. Pd: Voltage levels balance, fault-tolerant ability	Simulation and experimental.

TABLE 1. (Continued.) Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
21	[21], 2004.	Modular-multilevel converter	A 2MW-17-level prototype of the new MMLC family is presented. Pd: Behaviour at steady state, transient conditions and fault conditions.	Simulation and experimental (Stratix FPGA), dSpace (ds1103). Application: Wide range of high voltage applications, traction converters, operating directly on the power line, and network inerties.
22	[22], Jan, 2015	A review; Modular Multilevel Converter	A general overview of the basics of operation of the MMC along with its control challenges are discussed. Pd: Modulation, design, control, and modelling of the MMC	Application: HVDC Systems, Variable-Speed Drives, Dynamic Braking Chopper, flexible ac transmission system
23	[23],2018	General review; MLI based on Main Submodules: Structural point of view	This paper presents five Main Submodules (SMs)to be used as the basic structures of MLIs and has widely reviewed almost all presented multilevel inverters in these manuscripts based on the proposed SMs with variety connections. Pd: Number of components, the ability to create inherent negative voltage, working in regeneration mode and using single DC source. Comparison details of asymmetrical MLI included.	Application: Wide range of power electronics applications.
24	[24], 2020.	Topological Review: voltage source MLI with reduced device count	Summarize the recently developed multilevel topologies with a reduced component count, based on their advantages, disadvantages, construction, and specific applications. Pd: A detailed comparison in terms of switch, diode, capacitor, inductor, transformer count was performed and systematically summarized in tables.	Giving guidelines to further improve the current multilevel topologies more efficiently and compactly.
25	[25], 2015.	A Review: MLI topologies:	Recently proposed multilevel inverter topologies with reduced power switch count are reviewed. Pd: A detailed comparison of recently proposed MLI topologies is presented	Application: High and low power applications.
26	[26], 2017.	High-level inverter topology	A new topology with a lower voltage rating component to improve the performance by remedying the mentioned disadvantages are discussed. Pd: DC, efficiency , Power loss, Voltage stress, load power factor. The proposed topology has been compared with other similar topologies.	Simulation and experimental (dSPACE). / The proposed topology can be operated with a different real-time environment with good performance.
27	[27], 2016.	A 9-level inverter is developed by stacking two-hybrid 5-level inverter.	Generation of higher number of voltage levels by stacking inverters of lower multilevel structures with low voltage devices for drives is presented. Pd: Speed control of induction motor, A 9-level inverter is compared with other existing 9-level topologies and tabulated.	Simulation and experimental. (FPGA,TMS320F28335). Application: Induction motor drives.
28	[28], 2020.	Switched Capacitor Converter (SCC)	RDC based Basic Cell (BC) of a novel Switched-Capacitor Converter (SCC) and its generalized structure has been proposed. The detailed analysis of capacitor selection procedure for 13 levels SCMLI is presented. Pd: Switching loss, reduced devices, voltage balance,	Simulation and experimental. (dSPACE 1104). / The proposed SCMLI is verified for asymmetric dc-source configuration with resistive load condition.

TABLE 1. (Continued.) Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
29	[29], 2019.	Asymmetrical MLI topology	<p>A new single-phase MLI topology has been proposed in this paper to reduce the number of switches in the circuit and obtain higher voltage level at the output.</p> <p>Pd: Reduced Switch Count, power loss.</p>	<p>Simulation(PLECS software)and experimental.(dSPACE 1104). /</p> <p>Application: MLI topology with multiple extension capabilities.</p>
30	[30], 2017.	Packed U-Cell (PUC), a new reduced-structure multilevel converter	<p>The issues associated with PUC are addressed and two approaches as remedy are presented.</p> <p>Pd: Reduce component count, total blocking voltage, switch ratings and extending its performance to high voltage applications.</p>	<p>Simulation and experimental (TMS320-F28335).</p> <p>Application: High and low voltage applications.</p>
31	[31], 2017.	Three phase five-level inverter topology	<p>An optimized topology of three-phase, MLI topology configuration for five or higher-level operation with reduced switch count is discussed.</p> <p>Pd: Reduced switching device count, conduction and switching losses, efficiency, THD.</p>	<p>Simulation and experimental (MS320LF28069).</p> <p>Application: Solar PV, drives applications.</p>
32	[32], 2019.	Five-level inverter topology	<p>In this paper, five-level inverter topology based on the stacked cell approach to MLIs with reducing the number of switches, flying capacitors to cut size, weight and costs while facilitating higher reliability, simpler wiring and lesser electromagnetic interference are discussed.</p> <p>Pd: Fault-tolerant operation, loss steering and common-mode elimination. Comparison of the proposed topology with existing methods and highlight its merits and features.</p>	<p>Simulation and experimental.</p> <p>Application: High reliability electric drives (induction motor drive)</p>
33	[33], 2019.	Single-phase symmetrical and Asymmetrical MLI topology.	<p>The presented topologies can generate 9-level output voltage in a symmetrical configuration, 13-level and 17-level in asymmetrical configuration with a single cell. A detailed comparison has been done of the presented topology with recently proposed topologies in terms of DC sources.</p> <p>Pd: RDC, THD, semiconductor switches, capacitor and total blocking voltage.</p>	<p>Simulation and experimental (dSPACE-1103).</p> <p>Application: High power/ high voltage applications with equal and unequal DC voltage sources.</p>
34	[34], 2017.	Single DC source nine-level switched-capacitor boost inverter topology.	<p>In this paper proposed boost inverter topology with nine level output voltage waveform using a single dc-source and two switched capacitors.</p> <p>The number of devices and cost are highlighted by comparing the recent and conventional inverter topologies.</p> <p>The total voltage stress of the proposed topology is lower and has a maximum efficiency of 98.25%.</p> <p>Pd: Step-up inverter, switched capacitor, reduce switch count, efficiency. Comparison of different PWM techniques is presented.</p>	<p>Simulation(PLECS) and experimental.</p> <p>Application: Low and medium voltage applications.</p>

TABLE 1. (Continued.) Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
35	[35], 2019.	Review paper; Reduced Switch(RS) MLI Topologies, Symmetrical H-bridge based RS MLI, Asymmetrical H-bridge based RS MLI, and Modified RS MLI topologies.	This review paper focuses on a number of recently developed MLIs used in various applications. Each topology has been reviewed carefully based on the number of switch count, number of dc sources used, PIV, TSV, and applications. Pd: Recently developed RS MLI topologies, comprehensive analysis and comparative evaluation. RDC, comparison based on RSS MLI, RSA MLI and RSM MLI topologies included. / Switching loss, THD	Simulation and experimental. Application: Renewable energy integration and Drives Application.
36	[36], 2014.	Cascaded MLI with series connection of H-bridge; basic units	In order to generate all voltage levels (even and odd) single-phase cascaded multilevel inverter based on H-bridge unit is proposed. Nine different algorithms are proposed to determine the magnitudes of dc-voltage sources and it compared to the conventional topologies. Pd: RDC, blocked voltage.	Simulation and experimental. (89C52 microcontroller by ATMEL). Application: Drive and control of electrical machines, connection of renewable sources, FACTS devices.
37	[37], 2015.	Symmetric multilevel converter	Design of symmetric multilevel converter to enhance the circuit's performance is presented . Pd: proposed inverter is compared to the conventional topologies, conventional cascaded inverter, semi cascaded inverter and cross-switched inverter.	Simulation(MATLAB/Simulink) and experimental (ATMega 64 microcontroller by ATMEL). Application: Suitable for medium-voltage applications and where a higher number of output levels are required.
38	[38], Oct, 2014	Symmetric, Asymmetric, and cascade switched-Diode multilevel converter	Symmetric, asymmetric, and cascade switched-diode multilevel converter are proposed, which can produce many levels with minimum number of power electronic switches, gate driver circuits, power diodes, and dc voltage sources. Pd: RDC,THD, efficiency, standing voltage on switches.	Simulation and experimental. Application: High-voltage applications.
39	[39], Mar, 2018.	Seven-level Pack U-Cell inverter	In this paper present single-phase Pack U-Cell (PUC) MLI. The output voltage has a higher amplitude than the maximum dc-link value used in the topology as a boost operation. Pd: THD, efficiency, Power losses.	Simulation and experimental (dSpace 1103). Application: Power Quality, Renewable Energy Conversion, Photovoltaic Applications.
40	[40], June 2013.	Cascaded MLI topology	This suggested topology requires less switches and relevant gate driver circuits realising the same level in output voltage compared with traditional cascaded inverter. Pd: RDC,symmetric, asymmetric structures, PIV and power losses.	Simulation (MATLAB/Simulink) and experimental (DSPIC30F4011). Validate the practicability of the proposed symmetric and asymmetric multilevel inverters, which can generate all voltage steps for a test case nine-level symmetric and 17-level asymmetric inverter.
41	[41], 2016.	An Envelope Type (E-Type) Module:	The proposed E-Type module which can generate 13 levels with reduced components, It can be used in high voltage high power applications with unequal DC sources. Pd: RDC, THD, Power losses.	Simulation(MATLAB/Simulink) and experimental (Microcontroller ATMEGA16). Application: High voltage high power applications with unequal DC sources.

TABLE 1. (Continued.) Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
42	[42], June, 2011	Single-phase seven-level Grid-connected inverter	A single-phase seven-level inverter for grid-connected photovoltaic systems, with a PWM control scheme is proposed in this paper. Pd: RDC, THD.	Simulation(MATLAB Simulink) and experimental (TMS320F2812). Application: Photovoltaic system; grid-connected PV inverters.
43	[43], April 2015.	Boost dc-link cascaded MLI	A structure of single-phase seven-level boost dc-link cascaded MLI (BDCLMLI) is proposed. Pd: Reduction of voltage stress across the inverter switches, RDC, Power loss, THD.	Simulation (MATLAB)and experimental. Application: Uninterrupted Power Supply.
44	[44], 2014.	SHE-PWM Cascaded MLI with adjustable dc-voltage levels control	New Multilevel Selective Harmonic Elimination Pulse Width Modulation (MSHE-PWM) technique based transformerless Static Synchronous Compensator (STATCOM) system employing cascaded H-bridge inverter configuration is discussed. Pd: efficiency, THD, Reactive power (VAR) compensation.	Simulation(Matlab/Simulink)and experimental(dSPACE DS1104). Application: Reactive power(VAR) compensation; STATCOM.
45	[45], 2014.	Buck-Boost MLI	The proposed inverter is derived from Buck-Boost converter, which is with the ability to buck/boost the input voltage. Pd: The current sharing with modulation strategy is discussed.	Simulation and experimental. Application: Wide range varied dc-link voltage, such as renewable energy generation.
46	[46], 2019.	Switched-Capacitor (SC) based (2n+1)-level single-phase inverter.	This paper presents a novel Switched-Capacitor (SC) based (2n+1)-level single-phase inverter with a reduced number of components and input dc-voltage supply. Pd: quasi-resonant switching, efficiency, comparative analysis, thermal modelling and design guidelines, THD, loss calculation.	Simulation(MATLAB-Simulink-PLECS) and experimental (sb-RIO GPIC). Application: 500 VA prototype (Level Step-up Single-Phase Inverter) are presented.
47	[47] , 2016	Single phase step-up five-level inverter	The proposed topology can realize the multilevel inversion with high step-up output voltage, simple structure and reduced number of power switches. Pd: RDC, THD, summarize comparisons of the popular five-level inverter topologies and the proposed inverter.	Simulation and experimental. Application: Photovoltaic(PV) systems
48	[48],Sept. 2019.	Single dc-source-based seven-level Boost inverter	This paper discussed about an effective circuit arrangement of an MLI that can increase the number of output voltage levels with a lesser number of device count. Pd: RDC, THD. Device count-based comparative analysis is presented.	Simulation (MATLAB/Simulink) and experimental (FPGA-NEXYS 4). Application: Suitable for electric vehicles as less number of series-connected battery cells are required.
49	[49], Jan. 2009.	Single-phase five-level Photovoltaic (PV)inverter topology	A single-phase MLI utilizes two reference signals and a carrier signal to generate PWM switching signals for PV application. Pd: THD, efficiency.	Simulation and experimental (DSP TMS320F2812). Application: Grid-connected PV systems

TABLE 1. (Continued.) Taxonomy of multilevel inverters.

Sl. No.	[Ref.], Year	Topology	Objective and Parameter discussed (Pd)	Simulation / Experiment / Suggestions / Application
50	[50], 2017.	Modified quasi-Z-source cascaded hybrid five-level inverter	<p>This paper proposes a single-phase modified quasi-Z-source (MqZS) hybrid three-level inverter. In addition, a single-phase modified quasi-Z-source cascaded hybrid five-level inverter (MqZS-CHI) is designed by connecting two three-level PWM switching cells in series for producing a nine-level output voltage.</p> <p>Pd: THD, efficiency, boost factor.</p>	<p>Simulation and experimental (TMS320F28335).</p> <p>Application: Recommended for high output voltage with lower THD applications.</p>
51	[51], Oct. 2011.	Z-source-based MLI with reduction of switches	<p>This study presents a new inverter topology based on a mixture of cascaded basic units and one H-bridge unit. The cascaded basic units produce positive and zero-voltage levels and at the same time suggested inverter obtains positive, zero- and negative voltage levels.</p> <p>Pd: THD, RDC.</p>	<p>Simulation (MATLAB/Simulink) and experimental.</p> <p>Application: DVR: Power quality.</p>
52	[52], 2019.	Integrated semi-double stage based MLI with voltage boosting scheme	<p>This paper proposes a single-phase, seven-level, transformerless inverter. The proposed configuration achieves voltage boosting using a non-isolated interleaved buck-boost converter, which is fused with the inverter configuration through two switched capacitors (SC).</p> <p>Pd: Stray capacitor voltage, the common-mode voltage (CMV), Leakage current minimization, Semi-double stage system, Voltage boosting, power transmission ratio, efficiency.</p>	<p>Simulation (MATLAB) and experimental.</p> <p>Application: Photovoltaic Systems.</p>
53	[53], 1998	Switch-Mode dc-to-ac inverter using buck converter	<p>A dc-to-ac inverter of bidirectional buck topology using non-linear robust controller is proposed to achieve dynamic robust performances for both resistive and reactive loads.</p> <p>Pd: The proposed inverters are immune to the large disturbances in input voltage and load current and the output voltage remains steady state and dynamic stable.</p>	<p>Simulation and experimental./</p> <p>Dynamic performance of the proposed inverter is validated for both resistive and reactive loads.</p>
54	[54], Aug. 2013.	40-kVA SiC JFET Inverter	<p>This paper describes the concept, design, construction, and experimental investigation of a 40-kVA inverter with silicon carbide junction field-effect transistors (JFETs). The inverter was designed to reach an efficiency exceeding 99.5%.</p> <p>Pd: Silicon carbide (SiC), Power losses, efficiency measurements.</p>	<p>Simulation and experimental.</p> <p>Application: High-efficiency inverter; exceeding 99.5%</p>

applied to the system. Since the AFE converter operates at high switching frequency ranges, it is more flexible to control.

7) The system exhibits high dynamic performance. Hence, the output voltage remains dynamically unchanged when subjected to large disturbances in supply voltage or load currents.

8) The dc-to-dc converter based topology is highly compatible for implementing closed-loop control system.

The main outcomes of this article are as follows.

- Design, analysis and performance of a three-phase infinite level inverter-driven induction motor is performed.

- This topology has been tested with a resistive load and found to possess very good quality voltage and current waveforms in terms of THD.
- The dc-voltage requirement for generating a fixed ac-voltage output is much less than that required by other similar topologies, which is validated by SPWM control.
- The third harmonic injection modulation scheme has also been performed using this inverter and found that the dc-source utilization can be improved further.
- The efficiency of the inverter has also been found to be better since only one switch per phase is operated at a high frequency. All the switches in conventional inverters are operated at high frequency.
- Scalar and vector control of induction motor has also been performed using this topology. It has been found that both the controls exhibits better dynamic performance with this topology. Moreover, the ILI has been found to impart better performance to an induction motor drive.
- In conventional inverter circuits, the voltage applied across the terminals of the motor would be of absolute discrete values like V_{dc} , $-V_{dc}$, $V_{dc}/2$, $-V_{dc}/2$, etc. Hence, the instantaneous error voltage between the applied voltage and the desired sinusoidal voltage would be large. This error voltage manifests as torque ripples in the motor, deteriorating the performance of the motor.
- In the buck converter-based topology presented in this paper, the applied voltage is near to sinusoid and hence the error in the applied voltage is almost negligible. This results in negligible torque pulsations in the motor, thereby obtaining better performance from the motor. Open ended windings are used so that all the three buck converters can be connected to the same voltage source.

Now-a-days power, electronic industries and their application fields are changing from traditional silicon (Si) power semiconductor devices to potentially superior and high frequency operable counterpart, i.e., Wide Band Gap (WBG) devices, like Silicon Carbide (SiC) and Gallium Nitride (GaN) [54]. With the advent of WBG materials, standard Si technology is being replaced by high-frequency switches. If the Si semiconductor devices used in ILI topology are replaced with WBG devices, then there will be more improvement in efficiency and performance. Thereby, the system becomes more compact and operable at high frequency ranges.

This paper is organized as follows. In section II, a brief description of the basic structure of the proposed three-phase infinite level inverter topology is discussed. Principle of operation of ILI is discussed in section III. In section IV, modes of operations. In section V, the mathematical model of the proposed topology is discussed. In section VI, design of ILI. Criteria for selection of components are discussed in section VII. The efficiency calculation is discussed in section VIII. In section IX, the third harmonic injection method is discussed. In section X, scalar and vector control of ILI fed induction motor are presented. Simulation results

are presented in section XI. In Section XII, the comparison of ILI with conventional H-bridge inverters and AFE-RDC-MLI topologies are presented. In section XIII, discusses about the experimental setup and results. Finally, section XIV, concludes the paper and XV, provides with future scope.

II. THREE-PHASE INFINITE LEVEL INVERTER TOPOLOGY

The basic structure of infinite level inverter is a buck converter followed by an H-bridge. The proposed topology has the sole objective of reducing the count of passive elements and power semiconductor components without any loss of power quality in power conversion. Meanwhile, it reduces the switching and conduction losses, size and control complexity of the circuit. Three individual ILI circuits are combined to obtain a three-phase ILI topology which is shown in Fig. 1. The proposed topology consists of one high-frequency operated switch for every buck circuit and four low-frequency operated switches for every H-bridge; hence, one inductor and one capacitor per phase. This topology has a high degree of dc-source voltage utilization and low voltage stress across the switches as compared to traditional two-level and other similar inverter topologies. Other attributes of this circuit are the absence of shoot-through issue, less reverse recovery loss and body diode conduction loss in semiconductor switches.

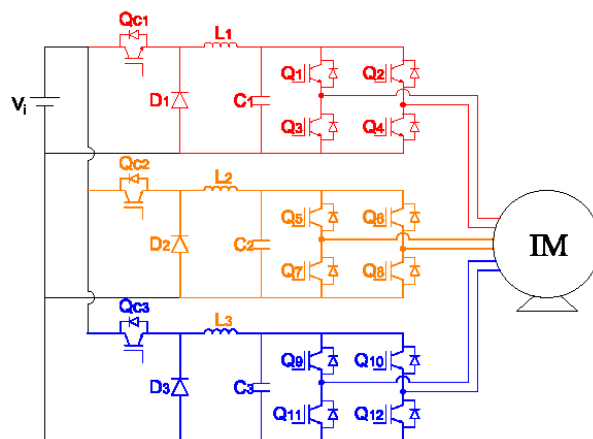


FIGURE 1. Three phase Infinite Level Inverter topology. Basic structure of the proposed topology is a buck converter (AFE converter) followed by an H-bridge. This topology consists of one high-frequency operated switch for every buck circuit and four low-frequency operated switches for every H-bridge; hence, one inductor and one capacitor per phase.

III. PRINCIPLE OF OPERATION

In an MLI each level is defined as the portion of voltage waveform where the magnitude of voltage remains constant for a defined duration. In the ILI, the voltage varies at every instant, and there is no time span where the voltage remains constant. If the time span in which the voltage remains constant tends to zero, the voltage waveform becomes a pure sine wave. The AFE converter generates a fully rectified infinite level voltage waveform by varying the duty cycle of the buck circuit in a fully rectified sinusoidal manner. The

quality of output voltage depends on the switching frequency of the AFE converter. Fig. 2 shows the SPWM control logic of ILI. The number of voltage levels developed across the buck capacitor is given by

$$V_{Level} = \frac{f_c}{f_m} \tag{1}$$

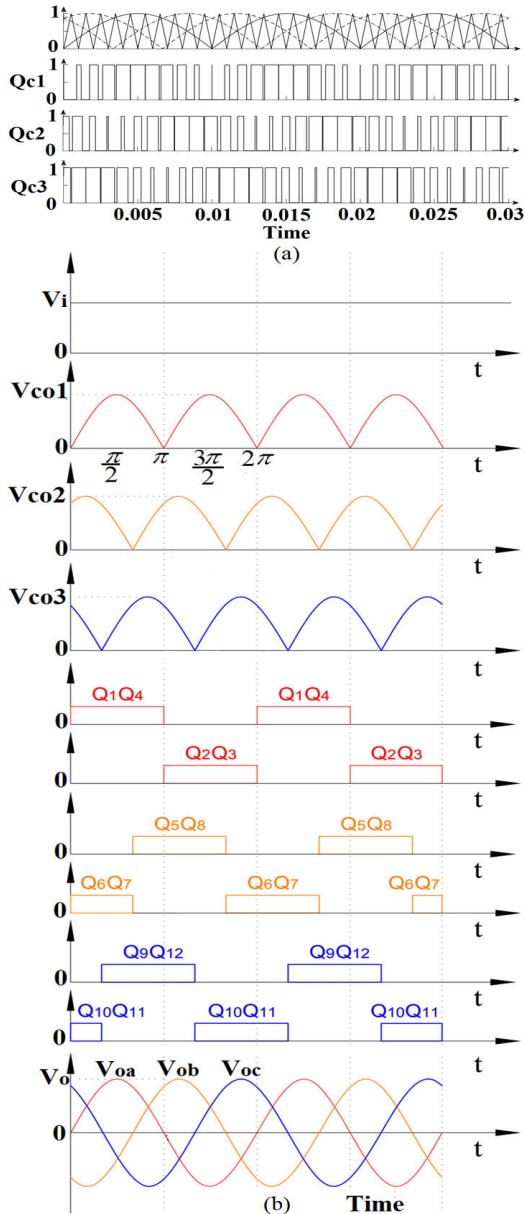


FIGURE 2. SPWM control logic of ILI. (a) High-frequency switching pulses of AFE converter (b) Power frequency switching patterns of polarity generation-part (H-bridge inverter).

where, ' f'_c ' is the carrier frequency and ' f'_m ' is the modulating signal. The duty cycle of the AFE converter is denoted by ' δ ' which varies in a rectified sinusoidal fashion.

$$\delta = M |\sin(\omega t)| \tag{2}$$

where, ' M ' is the modulation index. The output voltage of the buck converter is given by,

$$V_{co} = \delta V_i \tag{3}$$

The fully rectified sinusoidal output voltage across ' C'_1 ' is given by,

$$V_{co1} = V_i M |\sin(\omega t)| \tag{4}$$

where, $\omega = 2\pi f$ and ' V'_i ' is the input dc-source voltage. The rectified sine wave is unfolded by the H-bridge circuit to get a sinusoidal voltage across the load (7). The switching pulses for the H-bridge are synchronised with the modulating signal of the AFE converter. Therefore, the output voltage becomes

$$+V_i M |\sin(\omega t)| \quad 0 \leq \omega t \leq \pi \tag{5}$$

$$-V_i M |\sin(\omega t)| \quad \pi \leq \omega t \leq 2\pi \tag{6}$$

$$V_o = V_i M \sin(\omega t) \quad 0 \leq \omega t \leq 2\pi \tag{7}$$

The inversion process occurs at the natural zero-crossing point of the output voltage, resulting in zero voltage switching (ZVS) and hence reduces the switching power loss. In a three-phase ILI, the duty ratios of dc-to-dc converter switches are displaced by 120° . The three-phase voltage generation across the buck converters are fully rectified sine waveforms (8), (9), (10).

$$V_{co1} = V_i M |\sin(\omega t)| \quad 0 \leq \omega t \leq 2\pi \tag{8}$$

$$V_{co2} = V_i M |\sin(\omega t - 120)| \quad 0 \leq \omega t \leq 2\pi \tag{9}$$

$$V_{co3} = V_i M |\sin(\omega t + 120)| \quad 0 \leq \omega t \leq 2\pi \tag{10}$$

These rectified voltage waveforms are unfolded into the sinusoidal waveforms by appropriate switching of H-bridge.

$$V_{oa} = V_i M \sin(\omega t) \quad 0 \leq \omega t \leq 2\pi \tag{11}$$

$$V_{ob} = V_i M \sin(\omega t - 120) \quad 0 \leq \omega t \leq 2\pi \tag{12}$$

$$V_{oc} = V_i M \sin(\omega t + 120) \quad 0 \leq \omega t \leq 2\pi \tag{13}$$

IV. MODES OF OPERATION

There are four modes of operation for this inverter. They are discussed with the help of Fig. 3 Mode-1: During the period δT , the switch Q_{c1} is turned ON, and the inductor current starts rising. The diode is reverse biased. The output capacitor is charged exponentially. Mode-2: During the period $(1 - \delta) T$, switch Q_{c1} is turned OFF, and the parallel diode starts conducting. The current through the inductor falls, and it freewheels through the diode. In modes 1&2, Q_1 and Q_4 are conducting, so that a positive voltage is obtained across the load. Hence it generates the output voltage as $+V_i M |\sin(\omega t)|$ across the load side. Mode-3: is same as mode-1, except for the fact that Q_2 and Q_3 are switched ON. Mode-4: is same as mode-2, except for the fact that Q_2 and Q_3 are switched ON. Hence it generates the output voltage as $-V_i M |\sin(\omega t)|$ across the load side. The output voltage developed across the inverter load terminal is $V_o = V_i M \sin(\omega t); 0 \leq \omega t \leq 2\pi$.

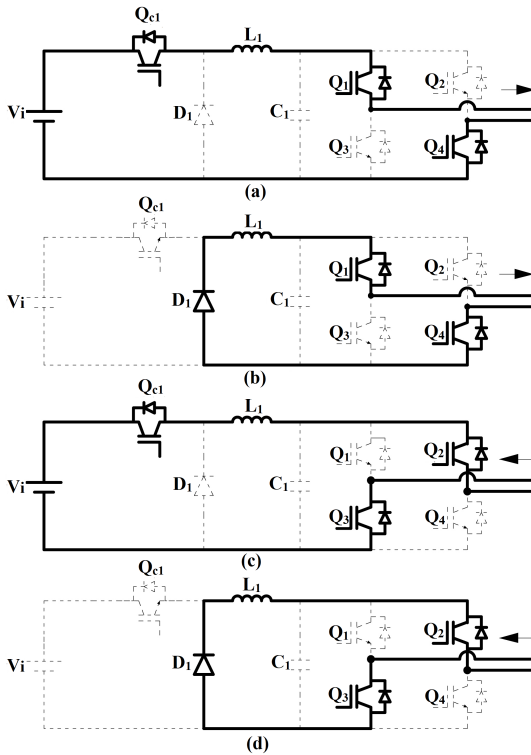


FIGURE 3. Mode of operations. (a) Mode-1, (b) Mode-2, (c) Mode-3, (d) Mode-4.

V. MATHEMATICAL MODEL OF ILI

The ILI is a buck–H bridge cascade. The buck converter is represented using the conventional transfer function of a dc-to-dc converter. The input to the transfer function is a rectified sine wave, acting as its variable duty cycle. The output for this transfer function is a rectified voltage, which would be available from the buck converter provided with such a duty cycle. This rectified voltage is unfolded with an H-bridge. This process is represented using a switch, which passes the input as such to the output when the input is positive and passes the negative of the input when the input is negative. This is equivalent to inversion process. The main part of the proposed converter is a dc-to-dc buck converter is shown in Fig.5. Here the buck converter plant is modelled to attain a simplified and linearized system around the equilibrium point using feedback control technique. The dynamic averaged equations of buck converter are Inductor current,

$$\frac{di_L}{dt} = \frac{-V_{CO}}{L} + \frac{V_d}{L} \tag{14}$$

Capacitor voltage,

$$\frac{dV_{CO}}{dt} = \frac{i_L}{C} - \frac{V_{CO}}{RC} \tag{15}$$

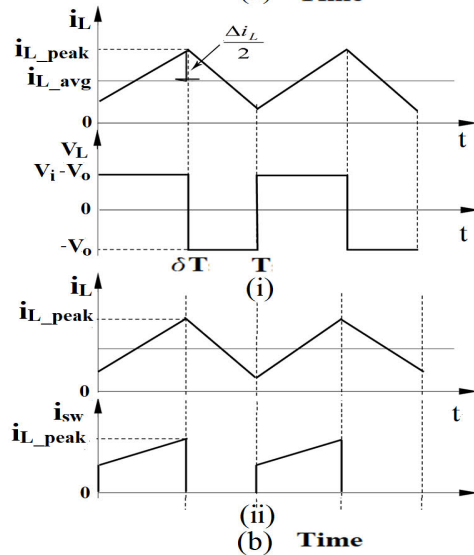
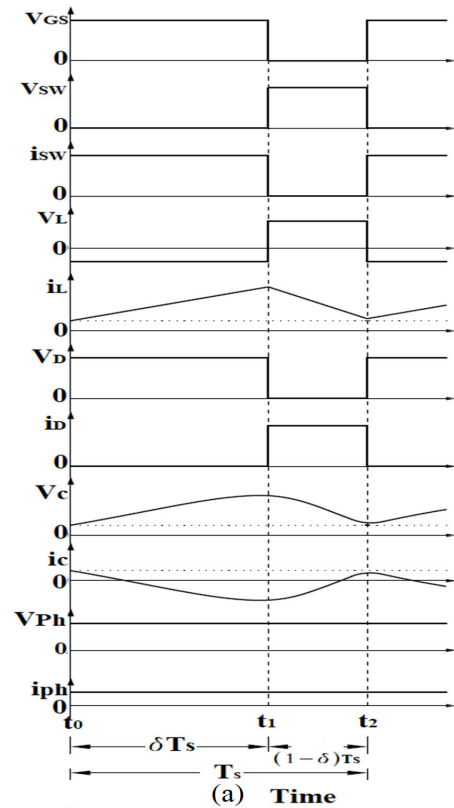


FIGURE 4. (a) AFE converter switching characteristics wave forms and (b) Waveforms of voltage and current of (i) Inductor and (ii) High-frequency switch.

On applying these equations into single input single output system model, the standard variables are used.

$$\left. \begin{aligned} \frac{dx_1}{dt} &= \frac{-1}{L}x_2 + \frac{V_i}{L}U \\ \frac{dx_2}{dt} &= \frac{1}{C}x_1 - \frac{1}{RC}x_2 \end{aligned} \right\} \tag{16}$$

where, system input 'U', system output 'y', and system state 'X'. On rewriting the system equations in terms of the

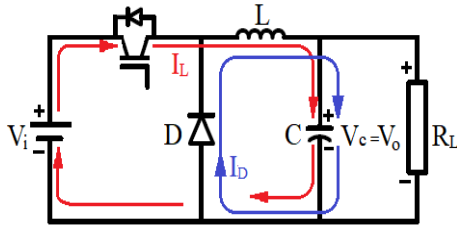


FIGURE 5. DC-to-dc buck converter topology.

standard variables, the inductor current, $i_L = x_1$, capacitor voltage, $V_{co} = x_2$ and duty ratio is the input 'u' and Output, $y = V_{co} = x_2$. Thereby the dynamic averaged equations can be rewritten with the new variables. Hence the state equations are (16). This equations satisfies linearity. By using sinusoidal PWM switching technique, the input dc-source voltage (V_i) is transformed into a fully rectified sinusoidal voltage waveform. Its frequency is same as the reference signal. In case of the proposed topology, the duty cycle is in a rectified sinusoidal fashion. Hence, the transfer function of input is

$$U(s) = \frac{1}{s^2 + 1} \tag{17}$$

To obtain the transfer function, from the state equation apply Laplace transform

$$\left. \begin{aligned} CY \left(S + \frac{1}{RC} \right) &= X_1 \\ SCY \left(S + \frac{1}{RC} \right) &= \frac{-1}{LC} Y + \frac{V_i}{CL} U \end{aligned} \right\} \tag{18}$$

On rearranging the equation,

$$Y \left(S^2 + \frac{1}{RC} S + \frac{1}{LC} \right) = \frac{V_i}{CL} U \tag{19}$$

The transfer function of the buck converter is given by the equation.

$$\frac{Y(s)}{U(s)} = \frac{V_i/LC}{S^2 + \frac{1}{RC} S + \frac{1}{LC}} \tag{20}$$

The rectified sinusoidal output voltage across buck capacitor is inverted using an H-bridge inverter. The converter output transfer function is mathematically expressed as

$$y(s) = \frac{V_i/LC}{S^2 + \frac{1}{RC} S + \frac{1}{LC}} \mathcal{L}(|\sin \omega t|) \tag{21}$$

For mathematical modelling, the following parameters are used. The design values are taken as $L = 9.7mH$, $C = 0.23\mu F$, $R_L = 50\Omega$. The numerator and denominator coefficients of the transfer function are $V_i/LC = 1519 \times 10^8$ and $s^2 + 8.7 \times 10^4 s + 4.5 \times 10^8$ respectively. Mathematical model of ILI and its output response are shown in Fig. 6.

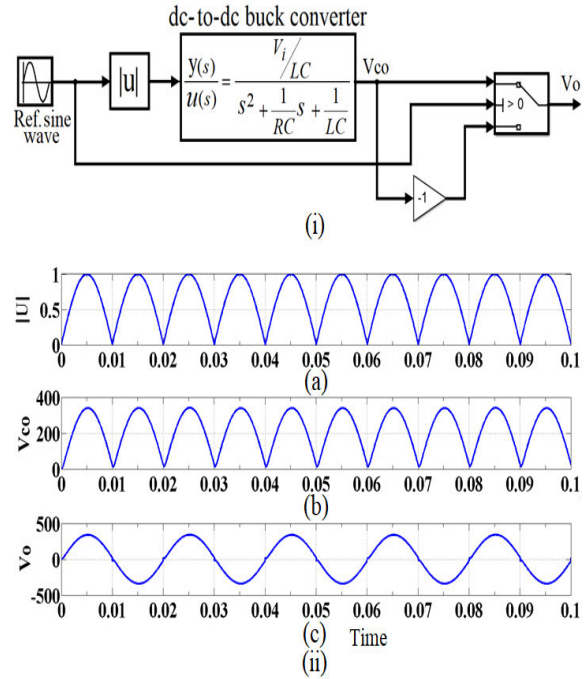


FIGURE 6. i) Mathematical model of proposed converter, (ii) Simulated output waveforms (a) reference signal, (b) rectified sinusoidal output voltage waveform generated by means of dc-to-dc buck converter transfer function, (c) sinusoidal output voltage waveform obtained from the inverter.

VI. DESIGN OF ILI

A. INDUCTOR (L)

In the ILI, the design of inductors plays a crucial role. The inductors are placed in the high-frequency operating part of the circuit, and it is fabricated using E65 ferrite core. When the switch 'Qc1' is turned ON for time period 'δT', the voltage across the inductor (V_L) is given by the difference between supply voltage and the output voltage.

$$V_L = L \frac{di_L}{dt} \tag{22}$$

The linear change in current (ΔI_L) through the inductor is,

$$\Delta I_L = \frac{V_i - V_o}{L} \delta T \tag{23}$$

Fig. 4(a) shows the voltage and current waveforms of the inductor and Fig. 4(b) shows those wave forms across the high frequency switch. The peak current through the inductor considered is $I_{Lavg} + \frac{\Delta I_L}{2}$. When the high frequency switch Q_{c1} is turned OFF, the freewheeling diode conducts during $(1 - \delta)T$ period and voltage of $-V_o$ appears across the inductor. Hence, voltage stress across the switch during this period is equal to V_i . The inductor current is controlled by two voltages which are appearing at both ends of the inductor (L). One is a high-frequency PWM pulsating voltage appearing across the diode (D) and the other is the voltage across the capacitor (C), which is fully rectified sinusoidal voltage waveform of 10 kHz. Output voltage is a pure sinusoid super imposed with ripple.

The average duty ratio is taken to be 50%. The rms value of the fundamental component of inductor current ripple ($I_{Lripple}$) can be calculated as

$$\frac{4}{\sqrt{2\pi}} \left[\frac{V_i}{2} \right] = (2\pi f_s L) I_{Lripple} \quad (24)$$

The duty cycle is getting varied continuously from 0 to δ_{max} in every half cycle of sinusoidal output wave. Depending upon the modulation index, δ_{max} will also vary. Hence, 50% duty cycle is taken as an average operating point. The maximum permitted amount of ripple is 5%. Therefore, $I_{Lripple} \leq 5\%I$ where I is rms value of the load current. The inductor value can be calculated as

$$L \geq \frac{\left(\frac{4}{\sqrt{2\pi}}\right)\left(\frac{V_i}{2}\right)}{(2\pi f_s) I_{Lripple}} \quad (25)$$

B. CAPACITOR (C)

The ripple in the load current is considered as negligible. Hence, the ripple current through the capacitor ($I_{Cripple}$) is equal to the ripple current of the inductor current. So, the ripple voltage of the capacitor is

$$V_{Cripple} = \frac{I_{Cripple}}{C\omega} \quad (26)$$

where $V_{Cripple} \leq 5\%V_o$ and in turn,

$$C \geq \frac{I_{Cripple}}{2\pi f_s (V_{Cripple})} \quad (27)$$

In order to ensure that output waveform is unaffected by the resonance caused by L and C , their values must be chosen such that

$$LC \leq \left(\frac{1}{2\pi 10f_s}\right)^2 \quad (28)$$

VII. SELECTION OF ILI COMPONENTS

A. SWITCHES

ILI has two different topological stages. (i) AFE Converter operating in power frequency and (ii) H-Bridge operating at high switching frequency.

Stage-I: In AFE converters, the high-frequency switches and diodes (D) are subjected to the same voltage. The voltage stress experienced across these devices is the same as the source voltage (V_i). During the turn OFF period $(1 - \delta)T$, the voltage stress across the high-frequency switch becomes equal to V_i , at the same time D is turned ON. During δT period, high frequency switches are turned ON and D is turned OFF. The diodes act as an open switch. Hence, the voltage stress across diode is V_i . In order to ensure safe operation of the switches, their ratings are taken to be higher than ' V_i '.

Stage-II: After every half cycle a couple of switches are switched ON, while the other switches are switched OFF. The effective voltage stress across the H-bridge switches is equal to the load voltage (V_o). Hence their voltage ratings must be higher than V_o . Current passes through the switches only when it is turned ON. The maximum current flowing

through the switches is equal to the peak current through the inductor (I_{Lpeak}).

$$I_{SWrating} = I_{Lpeak} \quad (29)$$

Voltage stress ($V_{SWstress}$) experienced across the high frequency switches, during the turn OFF period is

$$V_{SWstress} = V_i \quad (30)$$

B. DIODE

Current passes through the diode only when the high frequency operated switch is turned OFF. Its rating is chosen as

$$I_{Drating} = I_{Lpeak} \quad (31)$$

When the diode becomes reverse biased the voltage stress exerted across it is equal to ' V_i '.

$$V_{Dstress} = V_i \quad (32)$$

C. OUTPUT CAPACITOR

The maximum voltage (V_{Cmax}), applied across the capacitor is equal to V_i . The capacitor carries only the ripple component of current. Hence,

$$I_{Cpeak} = \Delta I_L \quad (33)$$

VIII. EFFICIENCY

Efficiency of any power electronic circuit is determined by the losses taking place in switches as well as in passive components. Losses in switching devices include switching losses and conduction losses. Switching loss depends on switching frequency and conduction loss depends on the load current. Since only one switch per phase is operated in high frequency, switching loss in this topology will be less. Moreover, total number of switches in this topology is much less than that in any other multilevel circuit, making conduction losses much lesser in this topology.

The simplified mathematical expression for the conduction and switching losses of IGBT switch in AFE converter is based on equation (34)

$$P_D = V_{ce_sat} I_m + \frac{1}{2} [(T_{ON} \frac{V_s I_m}{2}) + (T_{OFF} \frac{V_s I_m}{2})] f_{sw} \quad (34)$$

where P_D is the switching power dissipation, I_m is the collector current passing through the switch, V_{ce_sat} is the on-state drop, f_{sw} is the switching frequency, T_{ON} is the turn-on period, T_{OFF} is the turn-off period and V_s is the standing voltage of the switch. In case of diodes, the forward conduction and switching losses are considered. Meanwhile, the inductor and capacitor loss of the AFE converter is calculated by equation (35). ESR is also considered for the calculation of capacitor loss.

$$\left\{ (V_f I_m) + [V_D I_{rr} t_{rr} f_{sw}] \right\} + \left\{ \left(I_{Lpeak} + \frac{I_{ripple}}{2} \right) R_L \right\} + \{ 0.1 I_L ESR \} \quad (35)$$

where V'_D is the voltage across the diode. In polarity generation part, the H-bridge inverter switches are operating in soft-switching manner at fundamental frequency. The conduction and switching losses of H-bridge inverter comes to

$$(V_{ce-sat} I_m) N_{sw} + \frac{1}{2} \left[\left(T_{ON} \frac{V_s I_m}{2} \right) + \left(T_{OFF} \frac{V_s I_m}{2} \right) \right] f_L N_{sw} \quad (36)$$

The calculated value of switching losses per phase for both high frequency operating switch and H-bridge switches are 1.26W and 0.025W respectively. In theoretical loss calculation, the efficiency of the inverter is found to be 98%.

IX. THIRD HARMONIC INJECTION METHOD

The SPWM control technique used in conventional VSI, can obtain sinusoidal output voltage, with less harmonic distortion, however the maximum obtained output fundamental amplitude, is only 78.54 %. while, $M = 1$ (Line voltage = $0.6123V_iM$). To expand the output fundamental amplitude beyond from this limit, the ‘ $M > 1$ ’ have to be taken more than one. To improve the dc-bus voltage utilization of the inverter, without entering the over modulation, the third harmonics injection logic can be used. The third harmonics injection in the leg voltage references leads to an increase in the linear modulation range of three-phase VSI, by reducing the peak of the leg reference voltages and hence the modulation index can be pushed beyond the value of one without entering over modulation. By injecting the third harmonics with amplitude of one sixth of the fundamental harmonics, the maximum of the fundamental can be increased by 15.47% as compared to the SPWM control scheme (Line voltage = $1.4142V_iM$) The same concept is applied to the proposed topology, consequently the linear modulation range is extended. Hence, this increases the maximum fundamental output voltage without moving into the over modulation region. Moreover the third harmonic component does not affect the output phase voltages. Fig. 7 shows the logic for implementing the third harmonic injection PWM. It helps to improve the dc-bus voltage utilization of ILI as well as the fundamental amplitude of output line-to-line voltage by 15.47%. Thereby, effectively improve the dc-bus voltage utilization. The duty ratio of an ILI is

$$\delta_{ref} = |d \sin(\omega t)| \quad (37)$$

A third harmonic component of $k \sin(3\omega t)$ is superimposed with the sinusoidally varying duty cycle. The optimum point obtained is at $d = 2/\sqrt{3}$ and $k=1/6$ without any over modulation, Hence the reference signal is modified as equation (38)

$$\delta_{ref} = |d \sin \omega t + k \sin(3\omega t)| \quad (38)$$

The instantaneous value of the output voltage across the converter capacitance is

$$V_{co_fund} = |1.1547V_m \sin(\omega t)| + |\frac{1}{6} V_m \sin(3\omega t)| \quad (39)$$

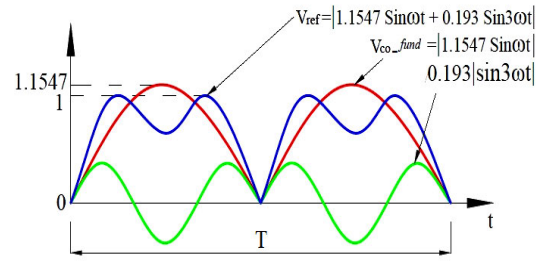


FIGURE 7. Third harmonic injection PWM control implementation logic.

The output line voltage across the load is

$$V_{o_fund} = 1.1547V_m \sin(\omega t) \quad (40)$$

The magnitude of three-phase output voltage becomes (41), (42), and (43).

$$V_{o1_fund} = 1.1547V_m \sin(\omega t); \quad 0 \leq \omega t \leq 2\pi \quad (41)$$

$$V_{o2_fund} = 1.1547V_m \sin(\omega t - 120); \quad 0 \leq \omega t \leq 2\pi \quad (42)$$

$$V_{o3_fund} = 1.1547V_m \sin(\omega t + 120); \quad 0 \leq \omega t \leq 2\pi \quad (43)$$

X. SCALAR AND VECTOR CONTROL OF ILI FED INDUCTION MOTOR

This section discusses about the speed control of open-ended three-phase induction motor using different control schemes, such as scalar (V/f) and direct vector control. Using the proposed ILI topology, the induction motor speed can be controlled with high dynamic performance.

A. SCALAR (V/F) CONTROL METHOD

The V/f control technique is an attractive method to control the induction machine speed because of its simplicity and user-friendly nature. Block diagram of scalar control implementation logic is shown in Fig. 8. Here, the air-gap flux of the induction machine can be controlled at desired value by proportionally varying stator voltages and frequency. Consequently, the machine retains its torque/ampere capacity at any speed. Meanwhile, the machine speed can be accurately maintained at any desired value under steady-state condition. However, at low-speed range, the torque capability is limited because of the voltage drop across the stator winding resistance, which is dominant.

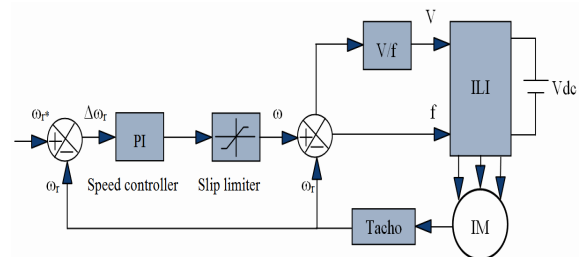
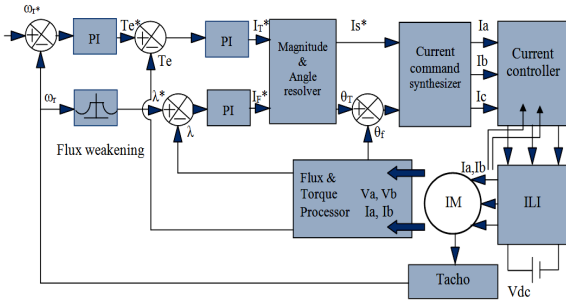


FIGURE 8. Scalar control of ILI fed induction motor implementation logic.

B. DIRECT VECTOR CONTROL METHOD

The block diagram of the direct vector control system is shown in Fig. 9. Here the field angle is calculated using


FIGURE 9. Direct vector control of ILI fed induction motor.

the terminal voltages and currents. The actual speed ω_r is compared with reference speed ω_r^* , the error is amplified and limited to generate the required reference torque T_e^* . The rotor flux linkage λ^* is kept at 1pu. Beyond 1pu, it is varied as a function of speed. This is to ensure that the rotor speed extends beyond the base speed by weakening the rotor flux linkages. The torque and flux references are compared to the actual torque T_e and flux linkage λ_r , calculated from the measured voltages and currents. The errors are amplified and limited to generate the required flux and torque producing components of current I_F^* and I_T^* . Phasor addition of these components yields the phasor reference current I_s^* and torque angle (θ_T) can be calculated from (52). The sum of torque angle and field angle gives the position of the stator current phasor I_s . Together with I_s^* , this generates the stator phase current references $I_{as}^*, I_{bs}^*, I_{cs}^*$.

The PWM control scheme is implemented using phase current control loops. The following equations represent the stator phase based calculation of various quantities.

$$V_{ds} = (R_s + L_s * p) I_{ds} + (L_m * p) I_{dr} \quad (44)$$

$$V_{qs} = (R_s + L_s * p) I_{qs} + (L_m * p) I_{qr} \quad (45)$$

$$I_{dr} = \frac{1}{L_m} \left\{ \int (V_{ds} - R_s I_{ds}) dt - L_s I_{ds} \right\} \quad (46)$$

$$I_{qr} = \frac{1}{L_m} \left\{ \int (V_{qs} - R_s I_{qs}) dt - L_s I_{qs} \right\} \quad (47)$$

$$\lambda_{dr} = L_r I_{qr} + L_m I_{qs} \quad (48)$$

$$\lambda_{qr} = L_r I_{dr} + L_m I_{ds} \quad (49)$$

Flux,

$$\lambda_r = \sqrt{\lambda_{dr}^2 + \lambda_{qr}^2} \quad (50)$$

Flux position,

$$\theta_f = \tan^{-1} \frac{\lambda_{qr}}{\lambda_{dr}} \quad (51)$$

$$\theta_T = \tan^{-1} \frac{I_T^*}{I_F^*} \quad (52)$$

Stator current phase angle $\theta = \theta_f + \theta_T$ Torque equation,

$$T_e = \frac{3P}{2} L_m (I_{qs}^* I_{dr} - I_{ds}^* I_{qr}) \quad (53)$$

Reference current is calculated as

$$I_s^* = \sqrt{(I_{ds}^*)^2 + (I_{qs}^*)^2} \quad (54)$$

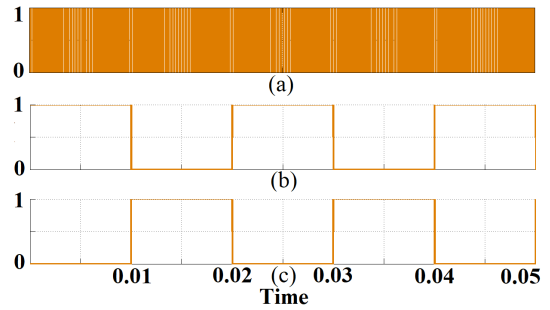
Three phase reference currents are

$$\left. \begin{aligned} I_{as}^* &= I_s^* \sin(\theta) \\ I_{bs}^* &= I_s^* \sin(\theta + 120) \\ I_{cs}^* &= I_s^* \sin(\theta - 120) \end{aligned} \right\} \quad (55)$$

XI. SIMULATION RESULTS OF ILI

A. SPWM CONTROL

The proposed inverter topology and its control are realized using MATLAB/Simulink environment. A three-phase balanced resistive load of 50Ω is taken for the simulation analysis. A switching frequency of 10 kHz is used for sine PWM pulse generation. Fig.10-13 shows the simulated waveforms of ILI, using resistive load. The power quality of the output voltage can be improved by increasing the AFE converter operating frequency. It significantly reduces the THD, which is less than the permissible limit as per IEEE-519 standards. The FFT analysis is shown in Fig. 14. This scheme requires only a 338V dc-source for obtaining 415V ac 50 Hz line-to-line output voltage.


FIGURE 10. Simulated waveforms of ILI. (a) High frequency, (b,c) Low frequency switching pulses.

B. THIRD HARMONIC INJECTION PWM CONTROL

The dc-source voltage utilization of the proposed inverter has been again validated using third harmonics injection PWM method. It requires only 293V dc-source voltage, in order to obtain 415V ac, 50Hz line-to-line output voltage. The simulated waveform of the third harmonic injection scheme is shown in Fig. 15. The proposed inverter is compared to a traditional two-level inverter topology. For a similar output voltage, a SPWM control scheme requires 654V dc-source voltage and an ordinary third harmonic injection scheme requires 564V dc-source. Fig.16 shows the comparison of dc-source voltage utilization between ILI and traditional inverter. The proposed topology has very high dc-source voltage utilization, compared to the traditional two-level inverter.

C. PERFORMANCE EVALUATION OF ILI TOPOLOGY USING THREE PHASE INDUCTION MOTOR

1) SPEED CONTROL USING SCALAR (V/F) CONTROL METHOD

The variable speed control performance of the proposed topology is validated using a three-phase induction motor. The speed control logic is implemented using constant V/f

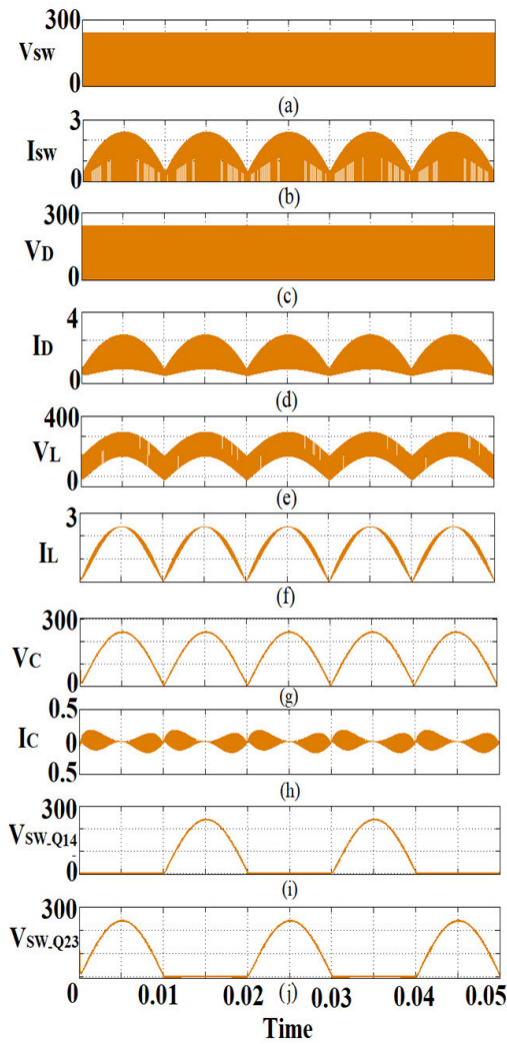


FIGURE 11. Simulated waveforms of ILI using resistive load. Voltage and current wave forms across the AFE converter components. (a,b) High frequency switch, (c,d) diode, (e,f) inductor, (g,h) capacitor, (i,j) voltage across low frequency operating switches.

scheme. The dynamic responses of the simulated output voltage waveforms using the V/f control are shown in Fig. 17. To verify the rotor start-up performance, the motor runs from standstill condition to a pre-defined speed of 300rad/s by applying 415V, 50Hz source voltage. Thereby, the rotor speed is attained and critically stabilized at that pre-defined speed with in a fraction of second.

2) SPEED CONTROL USING DIRECT VECTOR CONTROL METHOD

Direct vector control operation was simulated as per the block diagram mentioned in Fig.9. Reference speed was increased as well as decreased. The dynamic response of output voltage waveforms as well as the voltage across the buck capacitor is given in Fig. 18 for acceleration as well as deceleration.

XII. COMPARISON OF ILI WITH CONVENTIONAL H-BRIDGE INVERTERS AND AFE-RDC-MLI TOPOLOGIES

TABLE-2 compares the active components, passive components and THD of the 2-level H-bridge, 3-level H- bridge,

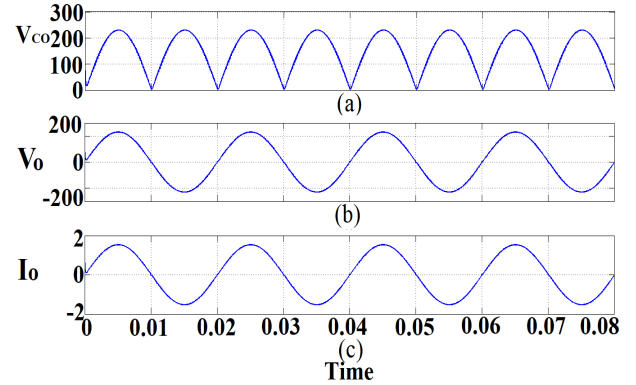


FIGURE 12. Simulated waveforms of ILI using resistive load. (a) Voltage waveform across the buck capacitor. (b) Voltage, (c) current waveforms across the load resistance.

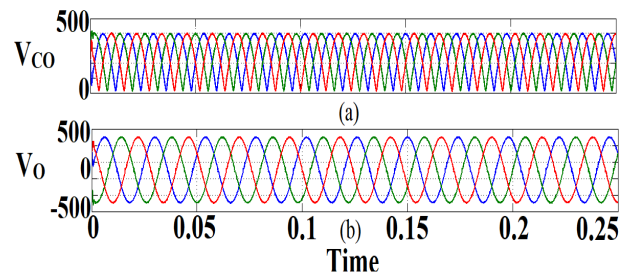


FIGURE 13. Simulated waveforms of ILI using resistive load. (a) Three-phase output voltage waveforms across the buck capacitor, (b) Three-phase output voltage wave form across the load resistance.

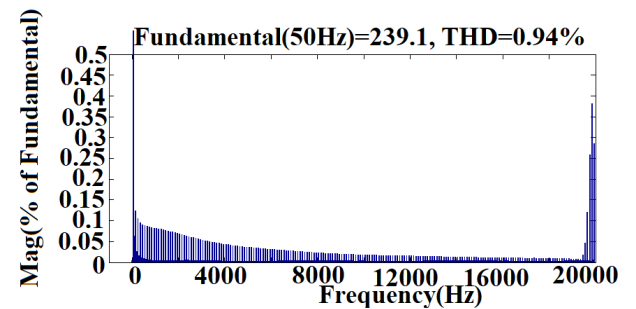


FIGURE 14. FFT analysis of output voltage waveform of the ILI.

5-level cascaded H-bridge MLI with the proposed ILI. Fig. 19 shows the simulation results of the output voltage wave forms of conventional and proposed inverter topologies. TABLE-4 shows the comparative analysis of different AFE-RDC-MLI (single phase circuit) topologies with proposed inverter. In comparison with conventional topologies, the proposed topology uses one high frequency operated switch and four power frequency operating switches per phase, thereby reduces the switching loss of the proposed inverter. Moreover, AFE converter circuit is a part of the proposed inverter that can generate non-finite number of voltage levels, which can improve the THD of the output voltage.

XIII. EXPERIMENTAL SETUP AND RESULTS

The hardware setup for both scalar and vector control implementation using ILI topology is shown in Fig. 20.

TABLE 2. Comparison between the proposed ILI and the conventional H-bridge inverters, (* Output voltage levels depends on switching frequency).

Sl. No.	Topology Type	High frequency operating switches (Sw)	Power frequency operating switches	Diode (D)	Inductor (L)	Capacitor (C)	dc-Source	Output voltage Levels	Switching frequency (fsw)	Control	THD of output voltage
1	H-Bridge (2-Level)	4	0	0	0	0	1	2	10 KHz	SPWM	58.34%
2	H-Bridge (3-Level)	4	0	0	0	0	1	3	10 kHz	SPWM	37.04%
3	Cascaded H-Bridge MLI (5-Level)	8	0	0	0	0	2	5	10 kHz	SPWM	35.38%
4	ILI (Proposed)	1	4	1	1	1	1	*	10 kHz	SPWM	1.20%

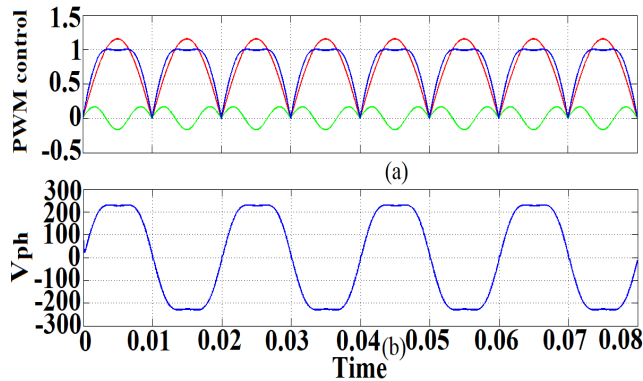


FIGURE 15. Simulated waveforms of (a) third harmonic injection PWM control implementation logic, (b) phase voltage waveform of the ILI using resistive load.

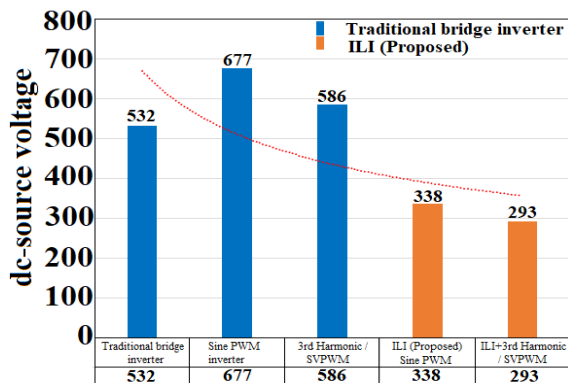


FIGURE 16. Comparison of dc-source voltage utilization of ILI with traditional bridge inverter.

TABLE-3 shows the parameters of ILI used for the experiment. For hardware tuning nearest available component values have been selected. Experiment was conducted on ILI topology under loads such as resistive and open-ended induction motor. Speed control of induction motor was carried out under constant V/f condition as well as under direct vector control method.

A. PERFORMANCE EVALUATION OF ILI TOPOLOGY USING SPWM CONTROL

Initially the performance of ILI topology was experimentally verified for a resistive load. NI PCIe-6351 card is used for data acquisition and control signal generation under

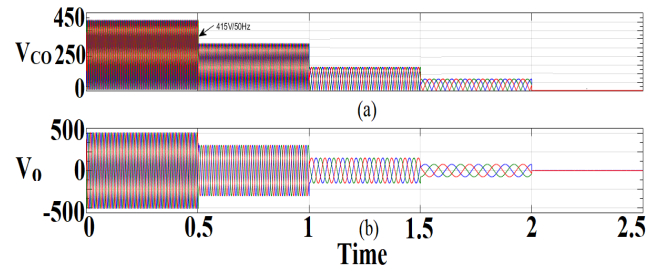


FIGURE 17. The dynamic responses of the simulated output voltage waveforms using V/f control. (a) Voltage waveform across the buck capacitor, (b) Line-to-line voltage across the load.

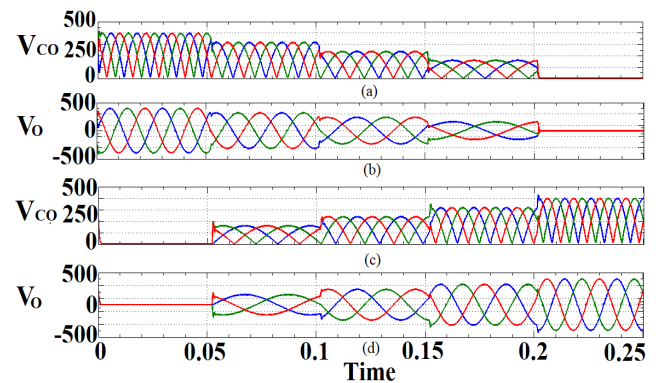


FIGURE 18. The dynamic responses of the simulated output voltage waveforms using direct vector control. (a,c) Voltage waveform across the buck capacitor, (b,d) Line-to-line voltage across the load.

TABLE 3. Parameters used for ILI.

Unit	Parameters	Values
kVA	Power rating	1
V	Inverter output voltage	240/415
Hz	frequency	50
mH	Inductor	9.7
μF	Capacitor	0.23
Ω	Load resistance	50
kHz	Switching frequency	10

MATLAB Real-Time Windows environment. SPWM control is implemented using this setup. The voltage level of the pulses generated by the PCI card is only 3.3V. In order to make this voltage sufficient enough to drive an IGBT, a voltage level shifter card is used as a buffer circuit. SEMIKRON IGBT modules- SKM50GB12T4 and MUR 860G ultrafast power diodes had been used for the hardware realization.

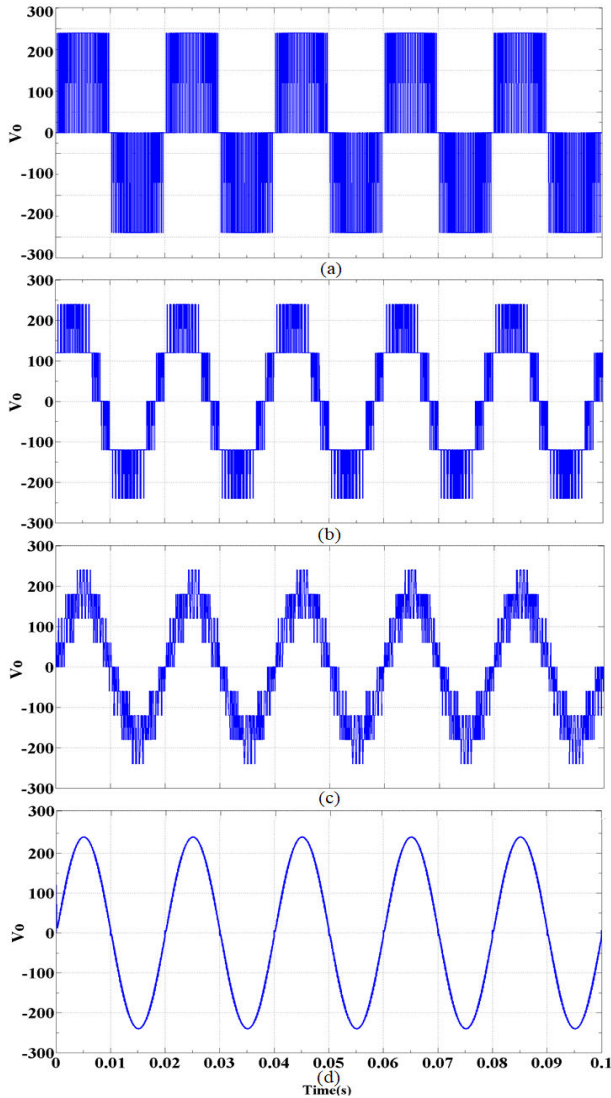


FIGURE 19. The simulated output voltage waveforms using resistive load (a) conventional 2-level H-bridge inverter, (b) 3-level H-bridge inverter, (c) 5-level cascaded H-bridge MLI, (d) Proposed topology.

Fully rectified sinusoidal output voltage is generated across the buck capacitor. This is then converted into sinusoidal voltage waveform using the H-bridge. The load voltage and current were measured using sensors LA 55-P and LV-25-1000, respectively. Fig. 21 shows the voltage and current waveforms of the AFE converter components. Fig. 22 shows the experimental results of ILI using resistive load and Fig. 23 shows the THD of ILI output voltage waveform using FLUK 434 energy analyser, and it is found to be 1.2%. The peak value of the pole voltage in a conventional inverter is $V_{dc}/2$. In case of ILI, the presence of the buck converter makes the peak value of per phase voltage equal to V_{dc} itself. In this case, this results in the reduction of dc-source requirement. The inherent filtering present with the buck converter makes the voltage a smoothly varying one, which results in the THD reduction.

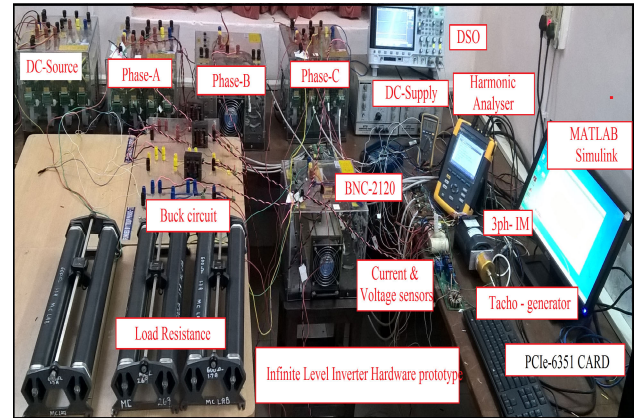


FIGURE 20. Experimental setup of an ILI to control an open-ended three-phase induction motor implemented using scalar and vector control logic.

B. PERFORMANCE EVALUATION OF ILI TOPOLOGY USING THIRD HARMONIC INJECTED PWM CONTROL

In order to increase the output voltage further without resorting to over modulation, harmonic injection control is used. Since triplen harmonic waves are in phase with all the three phase voltages, they cancel out at the line to line voltages. This property makes it possible to have a modulation index higher than one, without any over modulation of the modulating signal with respect to the carrier wave. DC-source voltage utilization of the proposed inverter is experimentally validated using third harmonic injection PWM method, which is almost similar to SVPWM technique. Fig. 24 shows the experimental voltage waveforms under this condition. The main advantage of third harmonic injection PWM is that, the fundamental amplitude of output line-to-line voltage increases by 15.4%, thereby effectively improving the dc-source voltage utilisation of the proposed MLI topology. The test results prove that, while it requires 338V dc-source voltage supply for developing 415V output voltage under SPWM control, third harmonic injected PWM control requires only 293V for maintaining the same voltage level.

C. PERFORMANCE EVALUATION OF ILI TOPOLOGY USING OPEN-ENDED THREE PHASE INDUCTION MOTOR

1) SPEED CONTROL USING V/f CONTROL METHOD

In case of conventional inverters, switching signals for V/f control are generated by comparing the modulation signal with a carrier wave. The two switches in the same leg are operated in a complimentary manner. In case of ILI, the modulation signal is rectified and compared with the carrier signal for generating switching signals for a buck converter. The practical testing of proposed ILI is carried out using a three-phase open-ended induction machine with NI-data accusation card prototyping system. Fig. 25 shows the block diagram of real-time implementation. The actual rotor speed of the induction motor is measured using electronic tacho-generator (analog output of 10V-dc at 1500 rpm),

TABLE 4. Comparative analysis of different AFE-RDC-MLI (single phase circuit) topologies with proposed inverter. (Output voltage levels depends on switching frequency).**

Sl. No.	Ref. No.	Topology Type	Sw	D	L	C	dc-source	AFE-converter	Output voltage levels	fsw	Polarity generation	Control	Efficiency	THD of output voltage	RDC	Applications
1	[49]	Conventional three-level PWM inverter	5	1	1	1	1	Boost	3	20kHz	H-Bridge	PWM	90%	8.00%	Yes	PV-Grid connected
2	[44]	Cascaded MLI	10	2	2	2	2	Buck	5	1.6kHz	H-Bridge	Multilevel Selective Harmonic Elimination -PWM	97%	23.45%	Yes	Power system , STATCOM
3	[45]	Buck-Boost MLI	8	0	2	1	1	Buck-Boost	5	24kHz	H-Bridge	Phase Opposition Disposition - PWM	-	22.65%	Yes	Renewable energy generation
4	[46]	Five-level inverter	8	1	1	3	1	Boost	5	20kHz	H-Bridge	Level Shifted - SPWM	96.50%	2.10%	Yes	For low voltage de-source
5	[47]	Step-Up Five-Level inverter.	6	3	1	2	1	Boost	5	15kHz	H-Bridge	Level Shift Multicarrier-PWM	-	4.80%	Yes	-
6	[50]	Modified quasi-Z-source cascaded hybrid inverter.	12	6	4	8	2	quasi-Z-source	5	10 kHz,	H-Bridge	Alternative phase opposition disposition	91%	3.60%	Yes	For high output voltage with lower THD
7	[49]	Five-level inverter	6	5	1	2	1	Boost	5	20kHz	H-Bridge	PWM	86%	5.40%	Yes	PV-Grid connected
8	[48]	Seven-level boost inverter.	8	4	1	2	1	Boost	7	5kHz	H-Bridge	Phase shifted -SPWM	-	14.92%	Yes	Electric vehicle
9	[42]	Seven-Level Grid-Connected Inverter	7	10	1	4	1	Boost	7	-	H-Bridge	PWM	Yes	3.90%	Yes	PV-Grid connected
10	[43]	Boost DC-link cascaded MLI	10	2	2	2	2	Boost	7	-	H-Bridge	Multicarrier-SPWM	-	4.16%	Yes	UPS and AC-Drive
11	[51]	Z-source-based MLI	10	3	6	6	3	Z source	7	-	H-Bridge	PWM	-	24.05%	Yes	Power quality; DVR
12	[52]	Integrated Semi-Double Stage based MLI	12	2	2	3	1	Interleaved -Boost	7	20 kHz,	H-Bridge	PWM	91.04%	1.26%	yes	PV-Grid connected
13	[38]	Twenty-Five-Level Cascade Switched-Diode Converter	12	4	0	0	4	Switched -Diode	25	-	H-Bridge	Fundamental frequency-switching	93.99%	2.19%	Yes	Renewable energy sources and medium -voltage applications
14	*	ILI	5	1	1	1	1	Buck	**	10kHz	H-Bridge	SPWM	98%	1.20%	Yes	Induction motor drive

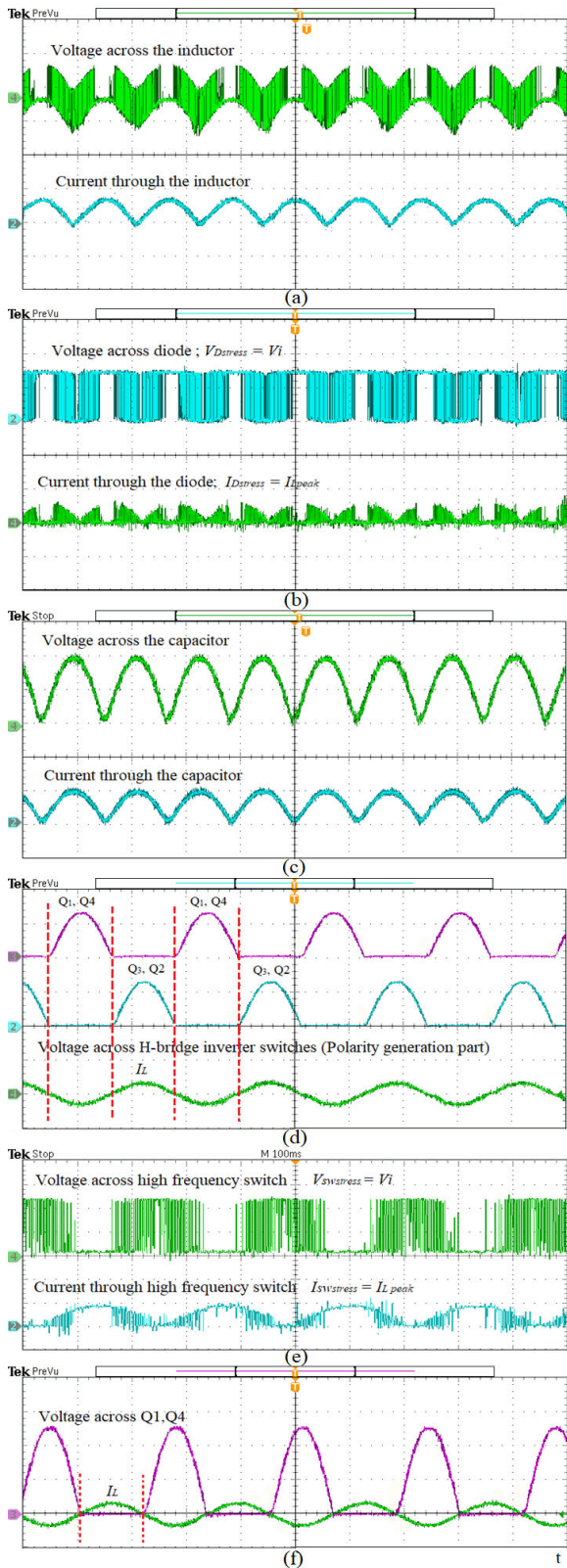


FIGURE 21. Hardware experimental results of ILI using resistive load. Voltage and current performance wave forms across the AFE Converter components (a) inductor, (b) diode, (c) capacitor, (d) voltage across low frequency operating switch, (e) voltage across high frequency operating switch, (f) Voltage and current waveform through Q1 and Q4.

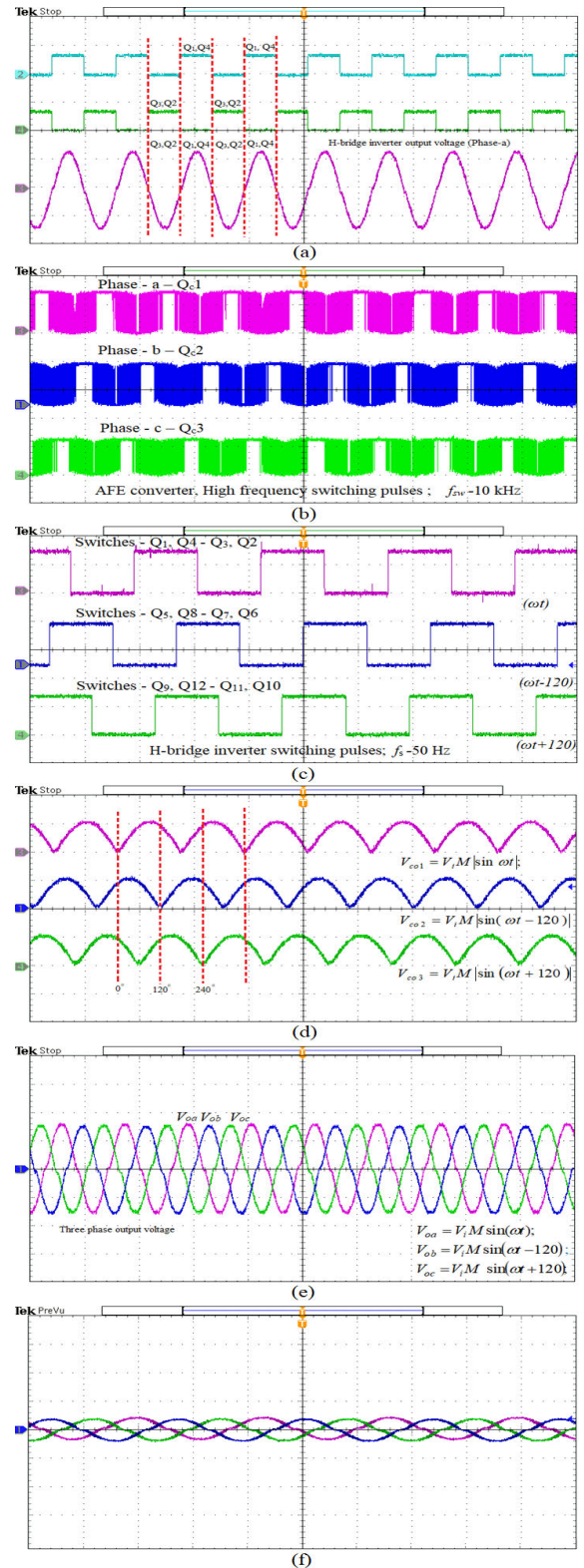


FIGURE 22. Hardware experimental results of three-phase ILI using resistive load. (a) H-bridge switching pulses and single-phase ILI output voltage waveform, (b, c) High and Low frequency switching pulses, (d) voltage waveforms across buck capacitors, (e) output voltage waveforms across the load terminals, (f) current waveforms.

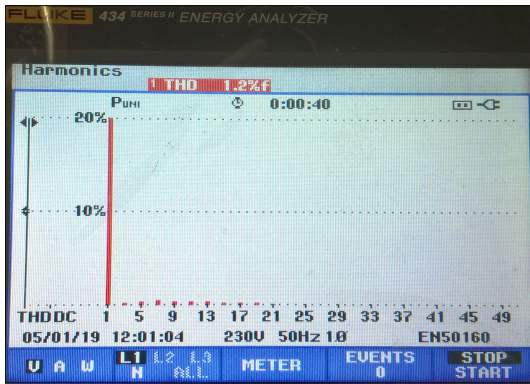


FIGURE 23. Hardware experimental result; THD of output voltage waveform of the ILI.

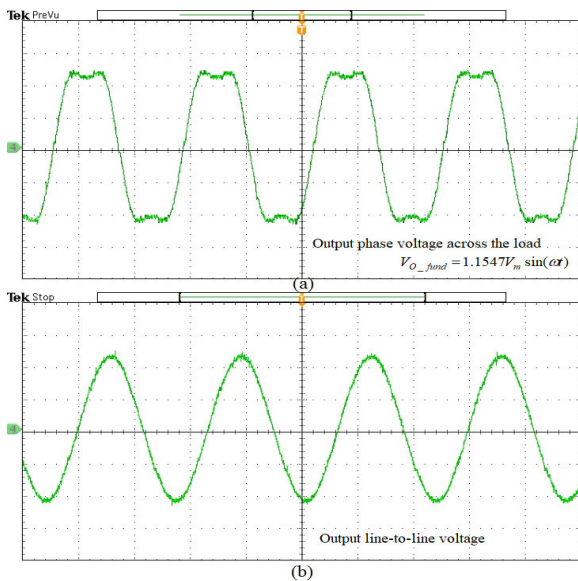


FIGURE 24. Hardware experimental results of ILI under resistive load, obtained by third harmonic injection PWM method. (a) Phase voltage waveform. (b) Line-to-line voltage waveform.

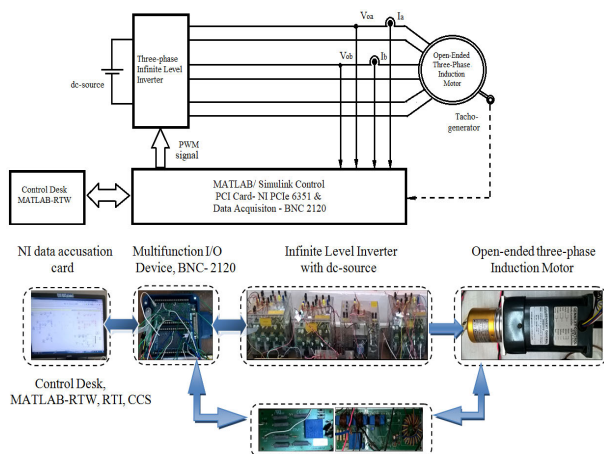


FIGURE 25. Real-time implementation block diagram.

and the corresponding dc-voltage value is fed into the computer through NI-PCIe-6351 Card. The capability of ILI for

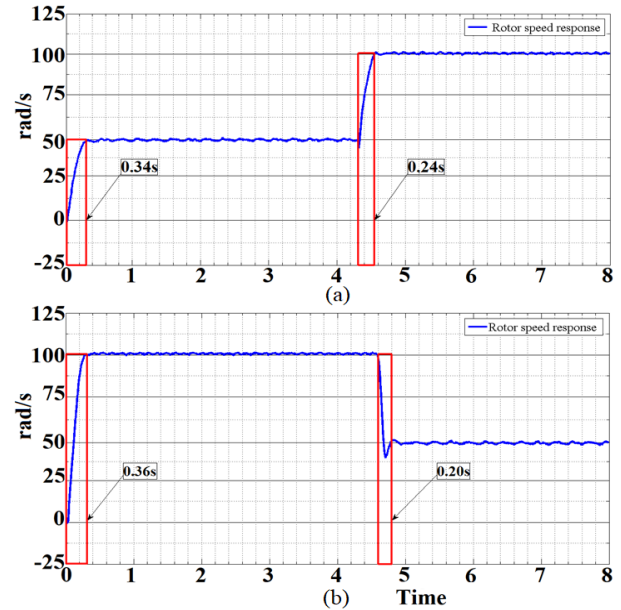


FIGURE 26. Experimental results of dynamic speed responses obtained from V/f control. (a) Increase of rotor speed from initial condition to pre-defined speed range 0 rad/s - 50 rad/s in 0.34s and then shifted to 100rad/s in 0.24s. (b) Rotor speed shifted from 0 rad/s to 100 rad/s in 0.36s and stepped-down speed transitions of varying rotor speed from 100 rad/s to 50 rad/s in 0.20s.

TABLE 5. Experimental data.

Motor speed responses	Speed control method V/f	
	Fig.26(a)	Fig.26(b)
Speed (Initial) in rad/sec.	0	0
Speed (Pre-fixed) in rad/sec.	50	100
Start-up time in sec.	0.34	0.36
Speed (Pre-fixed) in rad/s.	100	50
Speed (transition time) in Sec.	0.24	0.2
	Speed control method FOC	
	Fig.27	Fig.28
Speed (Initial) In rad/sec.	0	0
Speed (Pre-fixed) in rad/sec.	50	75
Start-up time in sec.	0.24	0.25
Speed (Pre-fixed) in rad/s.	100	135
Speed (transition time) in Sec.	0.12	0.12
Speed (Pre-fixed) in rad/sec.	50	75
Speed (transition time) in Sec.	0.11	0.12

running the induction machine under variable speed condition using V/f control scheme was experimentally verified. 10 kHz carrier signal was used for SPWM signal generation. In order to verify the rotor start-up performance, the motor was made to run from standstill condition to different pre-defined speed ranges. Fig. 26. (a) and (b) shows the MATLAB plot of speed responses of the induction motor. Rotor speed increased from 0 to 50 rad/s in 0.34s. It was further increased to 100 rad/s in 0.24s and got reduced to 50 rad/s in 0.2s.

2) SPEED CONTROL USING DIRECT VECTOR CONTROL METHOD

The induction motor speed can be more accurately controlled using the direct vector control logic. This scheme was implemented using the MATLAB/Simulink environment

TABLE 6. Comparison of simulation and experimental results.

Particulars	Simulation results	Figure Nos.	Hardware results	Figure Nos.	Parameters
Mathematical model of proposed converter. Fig. 6.	Simulated waveforms of ILI using resistive load. (a) Voltage waveform across the buck capacitor. (b) Voltage waveforms across the load resistance.	12(a,b)	Hardware experimental results using resistive load. (a) single-phase ILI output voltage waveform	22 (a)	Design values $L = 9.7mH$, $C = 0.23\mu F$, $R_L=50 \Omega$. $V_o = 240V$ $f = 50Hz$, $V_i = 338V$
Different switching pulses of ILI.	(a) High frequency, (b,c) Low frequency switching pulses	10 (a,b,c)	High and Low frequency switching pulses,	22 (b,c)	Switching frequency at High: $10kHz$, Low: $50Hz$.
Voltage and current wave forms across the AFE converter components. (Under resistive load)	(a, b) High frequency switch, (c, d) diode, (e, f) inductor, (g, h) capacitor, (i, j) voltage across low frequency operating switches.	11(a-j)	(a) inductor, (b) diode, (c) capacitor, (d) voltage across low frequency operating switch, (e) voltage across high frequency operating switch, (f) Voltage and current waveform through low frequency operating switch	21 (a-f)	The voltage stress across: High-frequency switches ($V_{SW\ stress} = V_i$) H-bridge switches = V_o . Diodes: ($V_{D\ stress} = V_i$) Inductors: = V_i , $V_o = 240/415V$
ILI output voltage waveforms across the load resistance using SPWM control.	(a) Voltage waveform across the buck capacitor. (b) Voltage and (c) current waveforms across the load resistance.	12(a,b,c)	(a) ILI output voltage (per phase) waveform.	22(a)	Output voltage(V_{ph}) = $240V$, $f = 50Hz$, $V_o = 1.4142V_i M$ $V_i = 338V$ for $V_o = 415V$ In traditional two-level inverter: • SPWM control scheme requires $654V$ dc-source voltage.
ILI three-phase output voltage waveforms (Under resistive load)	(a) Three-phase output voltage waveforms across the buck capacitor, (b) Three-phase output voltage waveform across the load resistance.	13 (a,b)	(d) Voltage waveforms across buck capacitors (e) Output voltage waveforms across the load terminals, (f) Current waveforms.	22(d-f)	ILI output voltage: $240/415V$
FFT analysis	FFT analysis of output voltage waveform of the ILI.	14	THD of output voltage waveform of the ILI	23	Simulation - 0.94% Hardware - 1.2%
i). Third harmonic injection PWM control (Under resistive load). ii). comparison of dc-source voltage utilization. (traditional two-level inverter v/s ILI)	i). (a) Third harmonic injection PWM control implementation logic. (b) Phase voltage waveform of the ILI. ii). dc-source voltage utilization between ILI and traditional inverter.	i).15(a,b) ii). 16	Third harmonic injection PWM method. (a) Phase voltage waveform. (b) Line-to-line voltage waveform	24(a,b)	i) $V_i = 293V$ for $V_o = 415V$ $V_o = 0.6123V_i M$ ii) In traditional two-level inverter: • An ordinary third harmonic injection scheme requires $564V$ dc-source voltage.
Dynamic speed responses obtained from V/f control. (Open-Ended Three-Phase Induction Motor load)	(a) Voltage waveform across the buck capacitor, (b) Line-to-line voltage across the load.	17(a,b)	(a) Increase of rotor speed from initial condition to pre-defined speed range. (b) Rotor speed shifted (retardation)	26(a,b)	(a) $0\ rad/s$ - $50\ rad/s$ in $0.34s$ and then shifted to $100rad/s$ in $0.24s$. (b) $0\ rad/s$ to $100\ rad/s$ in $0.36s$ and stepped-down speed transitions of varying rotor speed from $100\ rad/s$ to $50\ rad/s$ in $0.20s$
Dynamic speed responses obtained from direct vector control. (Open-Ended Three-Phase Induction Motor load)	(a, c) Voltage waveform across the buck capacitor, (b, d) Line-to-line voltage across the load.	18(a-d)	i). Increase of rotor speed from initial condition to a pre-defined speed ii). Increase of rotor speed from initial condition to a pre-defined speed. iii). (a, b). a.(i). Phase current and b.(ii). phase voltage across the motor winding.	i). 27 ii). 28 iii). 29	i). $0\ rad/s$ to $50\ rad/s$ in $0.24s$, then shifted to $100rad/s$ in $0.12s$ finally stepped-down to $50\ rad/s$ in $0.11s$. ii). $0\ rad/s$ to $75\ rad/s$ in $0.25s$, then shifted to $135\ rad/s$ in $0.12s$ finally stepped-down to $75\ rad/s$ in $0.12s$. iii). The rotor speed shifted from $130\ rad/s$ to $95\ rad/s$ by changing the phase voltage and frequency from $210V$, $45Hz$ to $150V$, $30Hz$ at $3.3s$. and vice versa.

with the same ILI hardware setup. The reference current components I_d and I_q obtained from flux and torque errors respectively are converted to reference phase voltages after subjecting them through axes transformations. These reference voltages form the modulation signals in the conventional

method. Here, these modulation signals are rectified and given as switching pulses to the buck converter. This rectified buck voltage is inverted using the H-bridge. Fig. 27 shows the dynamic speed responses obtained from the tacho-generator output of ILI fed induction motor. In order to validate the rotor

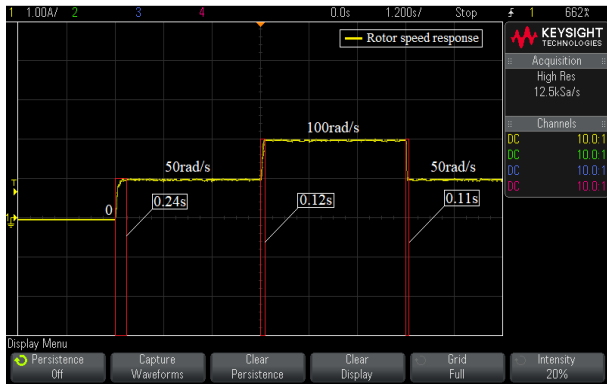


FIGURE 27. Experimental results of dynamic speed responses obtained from direct vector control. Increase of rotor speed from initial condition to a pre-defined speed of 0 rad/s to 50 rad/s in 0.24s, then shifted to 100rad/s in 0.12s finally stepped-down to 50 rad/s in 0.11s.

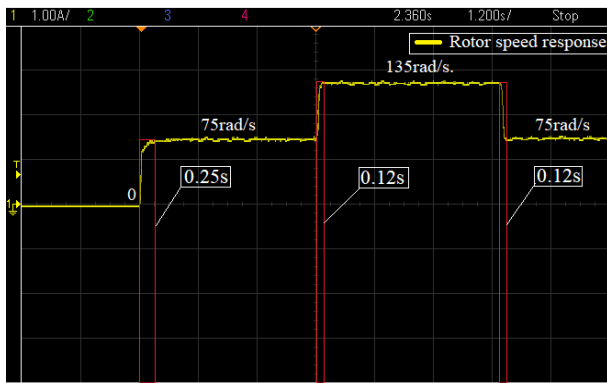


FIGURE 28. Experimental results of dynamic speed responses obtained from direct vector control. Increase of rotor speed from initial condition to a pre-defined speed of 0 rad/s to 75 rad/s in 0.25s, then shifted to 135 rad/s in 0.12s finally stepped-down to 75 rad/s in 0.12s.

start-up performance and acceleration, the motor was made to start from standstill condition to a pre-defined reference speed value of 50 rad/s. The rotor quickly attained the reference speed and critically stabilized in 0.24s. Again, the rotor was accelerated by changing the reference speed value to 100 rad/s. The rotor attained this speed in 0.12s. To test the retardation capability, reference speed was changed to 50 rad/s and the motor attained this speed in 0.11s. The step-up and step-down speed transition experiments were also conducted at various speed ranges starting from standstill condition to 75 rad/s in 0.25s, 135 rad/s in 0.12s and retarded back to 75 rad/s in 0.12s, as shown in Fig. 28. The transient and steady-state closed-loop performance of the inverter system over a wide range of speed is observed. Fig. 29.(a) shows the dynamic response of the phase current and voltage waveforms during retardation and Fig. 29.(b) shows the dynamic response of the phase current and voltage waveforms during acceleration. Here, the sensor conversion scales are 1:1 and 1:222 for current and voltage measurements, respectively. To validate the dynamic response and quality of power conversion of the ILI, the rotor speed is critically changed from 130 rad/s to 95 rad/s. It was repeated in reverse direction too.

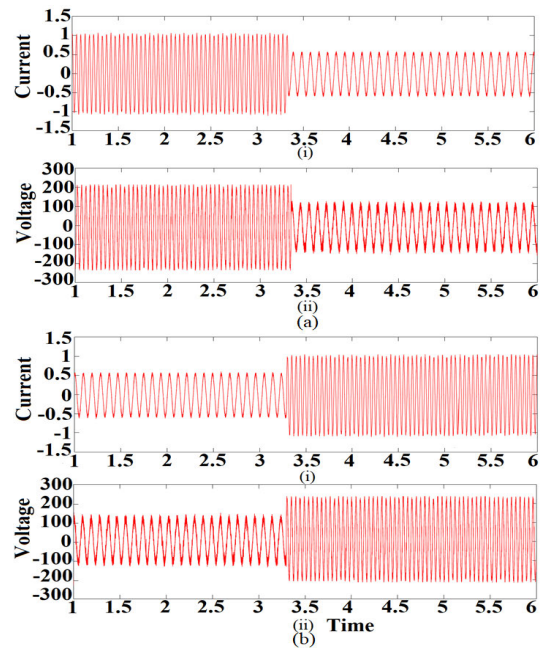


FIGURE 29. (a, b) The dynamic response of ILI fed induction motor obtained from direct vector control. (i) Phase current and (ii) phase voltage across motor winding. The rotor speed shifted from 130 rad/s to 95 rad/s by changing the phase voltage and frequency from 210V, 45Hz to 150V, 30Hz at 3.3s. and the vice versa. (Current and voltage sensor conversion scale is 1:1 and 1:222 respectively).

In TABLE-5 illustrate the experimental data of dynamic speed response and settling time of three-phase open-ended induction motor using ILI topology.

In comparison to scalar control technique the direct vector control exhibits an improved performance in settling time for speed responses. The experimental results prove that the quality of voltage and current waveforms remain constant before and after the speed transitions. Comparison of simulation and experimental results are summarised and presented in the TABLE- 6.

XIV. CONCLUSION

Design and analysis of the performance of an infinite level inverter driven induction motor have been discussed in this paper. ILI has been found to impart better performance to an induction motor drive. The ILI which belongs to an AFE-RDC-MLI topology has been tested with a resistive load and found to possess very good quality voltage and current waveforms in terms of THD. While conventional inverter topologies contain tens of percentage of THD, the topology mentioned in this paper contains a THD as low as 1.2%. Moreover, the dc- voltage requirement for generating a fixed ac-voltage output has been found to be much less than that required by other similar topologies, making the dc-source utilization better with this topology. While it is required to have a dc-voltage requirement of 677V in a conventional inverter working in sine PWM mode, the requirement of dc-voltage in the new inverter is only 338V.

Use of third harmonic injection modulation scheme has also been performed using this inverter and found that the

dc-source utilization can be improved further. Efficiency of inverter has also been found to be better, since only one switch per phase is operated at high frequency. All the switches in conventional inverters are operated at high frequency. Scalar and vector control of induction motor have also been performed using this topology. It has been found that the dynamic performance is better with this topology. This has been validated by accelerating and decelerating the machine with different reference speeds. Since the harmonic content in current has been very less, torque pulsations experienced by the motor would be negligible. Requirement of de-rating associated with induction motors fed by conventional inverters is not present in this case. Since there is no shoot-through menace, the chances of the inverter getting damaged is less, which results in better life and reliability of the drive system.

XV. FUTURE SCOPE

Silicon based power electronic devices are getting replaced with Wide Band Gap (WBG) devices. WBG devices are capable of switching at frequencies in the range of several megahertz. ILI can be a perfect fit for WBG devices, since the output voltage waveform can approach a pure sinusoidal status at these switching frequency levels. The performance of the drive system with advanced control schemes can also be studied.

APPENDIX

A. DESIGN OF INDUCTOR (L)

The rms value of the fundamental component of current ripple ($I_{Lripple}$) is obtained from,

$$\frac{4}{\sqrt{2\pi}} \left[\frac{V_i}{2} \right] = (2\pi f_s L) I_{Lripple} \quad (56)$$

where, V_i is the input voltage, $I_{Lripple}$ is the inductor current ripple, f_s is the switching frequency. The maximum permitted amount of ripple is considered as 5%. Therefore, ($I_{Lripple} \leq 5\%I$), Where, 'I' is the rms value of load current. Hence, the value of desirable inductor (L) is

$$L \geq \frac{\left(\frac{4}{\sqrt{2\pi}}\right)\left(\frac{V_i}{2}\right)}{2\pi f_s (0.05I)} \quad (57)$$

B. DESIGN OF CAPACITOR (C)

For designing the value of capacitor, assume the value of ripple in the load current as very small and hence, can be negligible. Here, the ripple current of the capacitor ($I_{Cripple}$) is equal to the ripple current of the inductor current ($I_{Lripple}$). The maximum permitted amount of the ripple voltage of the capacitor is considered as 5%. Therefore, $V_{Cripple} \leq 5\%V_o$.

$$V_{Cripple} = \frac{I_{Cripple}}{2\pi f_s C} \leq 0.05V_o \quad (58)$$

On rearranging the above equation, the value of the capacitor (C) is

$$C \geq \frac{I_{Cripple}}{2\pi f_s (0.05V_o)} \quad (59)$$

REFERENCES

- [1] P. Omer, J. Kumar, and B. S. Surjan, "A review on reduced switch count multilevel inverter topologies," *IEEE Access*, vol. 8, pp. 22281–22302, Jan. 2020.
- [2] J. Rodríguez, J.-S. Lai, and F. Z. Peng, "Multilevel inverters: A survey of topologies, controls, and applications," *IEEE Trans. Ind. Electron.*, vol. 49, no. 4, pp. 724–738, Aug. 2002.
- [3] L. M. Tolbert, F. Z. Peng, and T. G. Habetler, "Multilevel converters for large electric drives," *IEEE Trans. Ind. Appl.*, vol. 35, no. 1, pp. 36–44, Jan./Feb. 1999.
- [4] J.-S. Lai and F. Z. Peng, "Multilevel converters—A new breed of power converters," *IEEE Trans. Ind. Appl.*, vol. 32, no. 3, pp. 509–517, May 1996.
- [5] S. Kouro, M. Malinowski, K. Gopakumar, J. Pou, L. G. Franquelo, B. Wu, J. Rodríguez, M. A. Pérez, and J. I. Leon, "Recent advances and industrial applications of multilevel converters," *IEEE Trans. Ind. Electron.*, vol. 57, no. 8, pp. 2553–2580, Aug. 2010.
- [6] M. Malinowski, K. Gopakumar, J. Rodríguez, and M. A. Pérez, "A survey on cascaded multilevel inverters," *IEEE Trans. Ind. Electron.*, vol. 57, no. 7, pp. 2197–2206, Jul. 2010.
- [7] L. G. Franquelo, J. Rodríguez, J. I. Leon, S. Kouro, R. Portillo, and M. A. M. Prats, "The age of multilevel converters arrives," *IEEE Ind. Electron. Mag.*, vol. 2, no. 2, pp. 28–39, Jun. 2008.
- [8] J. Rodríguez, S. Bernet, B. Wu, J. O. Pontt, and S. Kouro, "Multilevel voltage-source-converter topologies for industrial medium-voltage drives," *IEEE Trans. Ind. Electron.*, vol. 54, no. 6, pp. 2930–2945, Dec. 2007.
- [9] L. A. Tolbert, F. Z. Peng, T. Cunnyngham, and J. N. Chiasson, "Charge balance control schemes for cascade multilevel converter in hybrid electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 49, no. 5, pp. 1058–1064, Oct. 2002.
- [10] X. Yuan and I. Barbi, "Fundamentals of a new diode clamping multilevel inverter," *IEEE Trans. Power Electron.*, vol. 15, no. 4, pp. 711–718, Jul. 2000.
- [11] M. F. Escalante, J. C. Vannier, and A. Arzandé, "Flying capacitor multilevel inverters and DTC motor drive applications," *IEEE Trans. Ind. Electron.*, vol. 49, no. 4, pp. 809–815, Aug. 2002.
- [12] S. R. Khasim, C. Dhanamjayulu, S. Padmanaban, J. B. Holm-Nielsen, and M. Mitolo, "A novel asymmetrical 21-level inverter for solar PV energy system with reduced switch count," *IEEE Access*, vol. 9, pp. 11761–11775, 2021.
- [13] D. J. Almkhles, J. S. M. Ali, S. Padmanaban, M. S. Bhaskar, U. Subramaniam, and R. Sakthivel, "An original hybrid multilevel DC-AC converter using single-double source unit for medium voltage applications: Hardware implementation and investigation," *IEEE Access*, vol. 8, pp. 71291–71301, 2020.
- [14] M. S. Bhaskar, D. Almkhles, S. Padmanaban, D. M. Ionel, F. Blaabjerg, J. He, and A. R. Kumar, "Investigation of a transistor clamped T-type multilevel H-bridge inverter with inverted double reference single carrier PWM technique for renewable energy applications," *IEEE Access*, vol. 8, pp. 161787–161804, 2020.
- [15] C. Dhanamjayulu, S. R. Khasim, S. Padmanaban, G. Arunkumar, J. B. Holm-Nielsen, and F. Blaabjerg, "Design and implementation of multilevel inverters for fuel cell energy conversion system," *IEEE Access*, vol. 8, pp. 183690–183707, 2020.
- [16] C. M. N. Mukundan, P. Jayaprakash, U. Subramaniam, and D. J. Almkhles, "Binary hybrid multilevel inverter-based grid integrated solar energy conversion system with damped SOGI control," *IEEE Access*, vol. 8, pp. 37214–37228, 2020.
- [17] S. Shuvo, E. Hossain, T. Islam, A. Akib, S. Padmanaban, and M. Z. R. Khan, "Design and hardware implementation considerations of modified multilevel cascaded H-bridge inverter for photovoltaic system," *IEEE Access*, vol. 7, pp. 16504–16524, 2019.
- [18] M. H. Mondol, M. R. Tür, S. P. Biswas, M. K. Hosain, S. Shuvo, and E. Hossain, "Compact three phase multilevel inverter for low and medium power photovoltaic systems," *IEEE Access*, vol. 8, pp. 60824–60837, 2020.
- [19] P. Ponnusamy, P. Sivaraman, D. J. Almkhles, S. Padmanaban, Z. Leonowicz, M. Alagu, and J. S. M. Ali, "A new multilevel inverter topology with reduced power components for domestic solar PV applications," *IEEE Access*, vol. 8, pp. 187483–187497, 2020.
- [20] A. Chen, L. Hu, L. Chen, Y. Deng, and X. He, "A multilevel converter topology with fault-tolerant ability," *IEEE Trans. Power Electron.*, vol. 20, no. 2, pp. 405–415, Mar. 2005.

- [21] M. Glinka, "Prototype of multiphase modular-multilevel-converter with 2 MW power rating and 17-level-output-voltage," in *Proc. IEEE 35th Annu. Power Electron. Spec. Conf.*, vol. 4, Dec. 2004, pp. 2572–2576.
- [22] S. Debnath, J. Qin, B. Bahrani, M. Saedifard, and P. Barbosa, "Operation, control, and applications of the modular multilevel converter: A review," *IEEE Trans. Power Electron.*, vol. 30, no. 1, pp. 37–53, Jan. 2015.
- [23] M. Vijeh, M. Rezanejad, E. Samadaei, and K. Bertilsson, "A general review of multilevel inverters based on main submodules: Structural point of view," *IEEE Trans. Power Electron.*, vol. 34, no. 10, pp. 9479–9502, Oct. 2019.
- [24] A. Salem, H. Van Khang, K. G. Robbersmyr, M. Norambuena, and J. Rodríguez, "Voltage source multilevel inverters with reduced device count: Topological review and novel comparative factors," *IEEE Trans. Power Electron.*, vol. 36, no. 3, pp. 2720–2747, Mar. 2021.
- [25] K. K. Gupta, A. Ranjan, P. Bhatnagar, L. K. Sahu, and S. Jain, "Multilevel inverter topologies with reduced device count: A review," *IEEE Trans. Power Electron.*, vol. 31, no. 1, pp. 135–151, Jan. 2016.
- [26] M. D. Siddique, S. Mekhilef, M. Rawa, A. Wahyudie, B. Chokaev, and I. Salamov, "Extended multilevel inverter topology with reduced switch count and voltage stress," *IEEE Access*, vol. 8, pp. 201835–201846, 2020.
- [27] R. V. Nair, S. A. Rahul, R. S. Kaarthik, A. Kshirsagar, and K. Gopakumar, "Generation of higher number of voltage levels by stacking inverters of lower multilevel structures with low voltage devices for drives," *IEEE Trans. Power Electron.*, vol. 32, no. 1, pp. 52–59, Jan. 2017.
- [28] T. Roy and P. K. Sadhu, "A step-up multilevel inverter topology using novel switched capacitor converters with reduced components," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 236–247, Jan. 2021.
- [29] M. D. Siddique, S. Mekhilef, N. M. Shah, A. Sarwar, A. Iqbal, M. Tayyab, and M. K. Ansari, "Low switching frequency based asymmetrical multilevel inverter topology with reduced switch count," *IEEE Access*, vol. 7, pp. 86374–86383, 2019.
- [30] A. N. Babadi, O. Salari, M. J. Mojibian, and M. T. Bina, "Modified multilevel inverters with reduced structures based on PackedU-cell," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 6, no. 2, pp. 874–887, Jun. 2017.
- [31] A. Hota, S. Jain, and V. Agarwal, "An improved three-phase five-level inverter topology with reduced number of switching power devices," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3296–3305, Apr. 2018.
- [32] A. Karthik and U. Loganathan, "A reduced component count five-level inverter topology for high reliability electric drives," *IEEE Trans. Power Electron.*, vol. 35, no. 1, pp. 725–732, Jan. 2020.
- [33] A. Chappa, S. Gupta, L. K. Sahu, S. P. Gautam, and K. K. Gupta, "Symmetrical and asymmetrical reduced device multilevel inverter topology," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 1, pp. 885–896, Feb. 2021.
- [34] M. D. Siddique, S. Mekhilef, N. M. Shah, N. Sandeep, J. S. M. Ali, A. Iqbal, M. Ahmed, S. S. M. Ghoneim, M. M. Al-Harthi, B. Alamri, F. A. Salem, and M. Orabi, "A single DC source nine-level switched-capacitor boost inverter topology with reduced switch count," *IEEE Access*, vol. 8, pp. 5840–5851, 2020.
- [35] P. R. Bana, K. P. Panda, R. T. Naayagi, P. Siano, and G. Panda, "Recently developed reduced switch multilevel inverter for renewable energy integration and drives application: Topologies, comprehensive analysis and comparative evaluation," *IEEE Access*, vol. 7, pp. 54888–54909, 2019.
- [36] E. Babaei, S. Laali, and S. Alilu, "Cascaded multilevel inverter with series connection of novel H-bridge basic units," *IEEE Trans. Ind. Electron.*, vol. 61, no. 12, pp. 6664–6671, Dec. 2014.
- [37] M. R. J. Osukee, E. Salary, and S. Najafi-Ravadanegh, "Creative design of symmetric multilevel converter to enhance the circuit's performance," *IET Power Electron.*, vol. 8, no. 1, pp. 96–102, 2015.
- [38] R. S. Alishah, D. Nazarpour, S. H. Hosseini, and M. Sabahi, "Novel topologies for symmetric, asymmetric, and cascade switched-diode multilevel converter with minimum number of power electronic components," *IEEE Trans. Ind. Electron.*, vol. 61, no. 10, pp. 5300–5310, Oct. 2014.
- [39] H. Vahedi, M. Sharifzadeh, and K. Al-Haddad, "Modified seven-level pack U-cell inverter for photovoltaic applications," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, vol. 6, no. 3, pp. 1508–1516, Sep. 2018.
- [40] A. Ajami, M. R. J. Osukee, A. Mokhberdoran, and A. van den Bossche, "Developed cascaded multilevel inverter topology to minimise the number of circuit devices and voltage stresses of switches," *IET Power Electron.*, vol. 7, no. 2, pp. 459–466, Feb. 2014.
- [41] E. Samadaei, S. A. Gholamian, A. Sheikholeslami, and J. Adabi, "An envelope type (E-type) module: Asymmetric multilevel inverters with reduced components," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7148–7156, Nov. 2016.
- [42] N. A. Rahim, K. Chaniago, and J. Selvaraj, "Single-phase seven-level grid-connected inverter for photovoltaic system," *IEEE Trans. Ind. Electron.*, vol. 58, no. 6, pp. 2435–2443, Jun. 2011.
- [43] R. Uthirasamy, U. S. Ragupathy, and V. K. Chinnaiyan, "Structure of boost DC-link cascaded multilevel inverter for uninterrupted power supply applications," *IET Power Electron.*, vol. 8, no. 11, pp. 2085–2096, Nov. 2015.
- [44] L. K. Haw, M. S. A. Dahidah, and H. A. F. Almurib, "SHE-PWM cascaded multilevel inverter with adjustable DC voltage levels control for STATCOM applications," *IEEE Trans. Power Electron.*, vol. 29, no. 12, pp. 6433–6444, Dec. 2014.
- [45] L. Wang, G. Ma, S. Zhao, and Y. Diao, "Study of a new buck-boost multilevel inverter," in *Proc. Int. Power Electron. Appl. Conf. Expo.*, Nov. 2014, pp. 78–81.
- [46] M. N. H. Khan, M. Forouzes, Y. P. Siwakoti, L. Li, and F. Blaabjerg, "Switched capacitor integrated (2n + 1)-level step-up single-phase inverter," *IEEE Trans. Power Electron.*, vol. 35, no. 8, pp. 8248–8260, Aug. 2020.
- [47] F. Gao, "An enhanced single-phase step-up five-level inverter," *IEEE Trans. Power Electron.*, vol. 31, no. 12, pp. 8024–8030, Dec. 2016.
- [48] N. K. Pilli, M. Raghuram, A. Kumar, and S. K. Singh, "Single DC-source-based seven-level boost inverter for electric vehicle application," *IET Power Electron.*, vol. 12, no. 13, pp. 3331–3339, Nov. 2019.
- [49] J. Selvaraj and N. A. Rahim, "Multilevel inverter for grid-connected PV system employing digital PI controller," *IEEE Trans. Ind. Electron.*, vol. 56, no. 1, pp. 149–158, Jan. 2009.
- [50] A.-V. Ho and T.-W. Chun, "Single-phase modified quasi-Z-source cascaded hybrid five-level inverter," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 5125–5134, Jun. 2018.
- [51] M. R. Banaei, R. Alizadeh, H. Khounjahan, E. Salary, and A. R. Dehghanzadeh, "Z-source-based multilevel inverter with reduction of switches," *IET Power Electron.*, vol. 5, no. 3, pp. 385–392, Mar. 2012.
- [52] S. Dhara and V. T. Somasekar, "An integrated semi-double stage-based multilevel inverter with voltage boosting scheme for photovoltaic systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 8, no. 3, pp. 2326–2339, Sep. 2020.
- [53] P. C. Sen and Z. Yang, "A new DC-to-AC inverter with dynamic robust performance," in *Proc. IEEE Region Int. Conf. Global Connectivity Energy, Comput., Commun. Control (TENCON)*, vol. 2, Dec. 1998, pp. 387–390.



His research interests include power electronics, drives, multilevel inverters, power quality, and power systems.



He had also functioned as the Chairman of the Board of Electrical Engineering streams of Calicut University and Kerala Technological University, Kerala. He had also associated with consultancy projects for industries. His research interests include power electronics, drives, neural network applications to power electronics, FACTS, and real time DSP implementations. He had received the Career Award for Young Teachers of AICTE, in 2000.

...



Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Dispersion analysis of nanofillers and its relationship to the properties of the nanocomposites

Gibin George^{a,*}, Amal P. Dev^b, N. Nikhil Asok^a, M.S. Anoop^b, S. Anandhan^{c,*}

^aDept. of Mechanical Engineering, SCMS School of Engineering and Technology, Pallissery, Ernakulam, Kerala, India

^bDept. of Automobile Engineering, SCMS School of Engineering and Technology, Pallissery, Ernakulam, Kerala, India

^cDept. of Metallurgical and Materials Engineering, National Institute of Technology Karnataka, Surathkal, Karnataka, India

ARTICLE INFO

Article history:

Received 3 March 2021

Received in revised form 8 May 2021

Accepted 12 May 2021

Available online xxxxx

Keywords:

Nanocomposite

Crystallization

Halloysite nanotube

Image processing

ABSTRACT

The dispersion and distribution characteristics of the reinforcements are the key reasons that influence the mechanical properties of the nanocomposites. In this paper, the dispersion and distribution analysis of nanofillers in a representative polymer is performed and the results are correlated to the crystalline and mechanical properties of the nanocomposite. The nanocomposite used in the present study is Elvaloy[®]4924 (EVACO)/halloysite nanotubes (HNTs) composite. The dispersion of halloysite nanotubes in the EVACO matrix is recorded as aluminum elemental maps obtained from energy dispersive spectroscopy (EDS). The dispersion and distribution of fillers in the composite are quantified using an image processing technique and it is correlated to the crystalline and tensile properties of the composites. The better dispersion and distribution of HNTs at 1wt.% filler loading resulted in a remarkable improvement in the crystallinity of the composite, which is measured by X-ray diffraction (XRD) and differential scanning calorimetry (DSC). The tensile strength was highest for composites loaded with 1 wt.% filler, and the strength decayed as the loading was further increased. Agglomeration of halloysite nanotubes and polymer-filler debonding was the major reason behind the reduction in tensile strength with filler loading, as observed in the scanning electron micrographs of the fractured surfaces.

© 2021 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the International Conference on Sustainable materials, Manufacturing and Renewable Technologies 2021.

1. Introduction

The addition of nanomaterials in polymer matrices is extensively used for refining the thermal, mechanical, flame resistance, and electrical characteristics of polymers. The improvement in these properties depends on the nature of the nanofiller used and their interaction with the respective polymer matrix. The naturally existing nanomaterials which are abundant, low cost, and non-toxic are used in the bulk production of polymer nanocomposites (PNCs) [1–3]. Clay and its minerals are the most commonly used nanofillers in PNCs because of their large availability, good interaction with the matrix, and ability to exfoliate into two-dimensional nanostructured layers [4].

The specific surface area of the filler and the interaction at the interface of the matrix and the filler have a vital role in improving the properties of the composite. In nanocomposites, the interaction between the polymer and the nanofiller is better than their micro-sized counterparts due to the large surface area of the nanoparticles [5]. The high surface area and associated high surface energy of the nanomaterials reduce the number of nanoparticles required to achieve a significant improvement in the properties of the composite [6]. As compared with the micro-sized particles, the quantity of nanoparticles required is only 1/100th to achieve the same properties in the composite [7]. The interaction of the fillers and polymer matrix in a polymer composite is a function of the surface area of the nanoparticles, and the nanoparticles have a large surface area as compared to the micro-sized counterparts of the same weight.

In polymer matrices, the addition of nanofillers exhibits a simultaneous enhancement in several properties of the matrix. Carbon nanotubes, for example, can simultaneously enhance polymer matrices' crystallization [8], tensile properties [9],

* Corresponding authors.

E-mail addresses: gibingeorge@scmsgroup.org (G. George), amaldev@scmsgroup.org (A.P. Dev), nikhil@scmsgroup.org (N.N. Asok), anoopms@scmsgroup.org (M.S. Anoop), anandtmg@gmail.com, anandhan@nitk.edu.in (S. Anandhan).

<https://doi.org/10.1016/j.matpr.2021.05.285>

2214-7853/© 2021 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the International Conference on Sustainable materials, Manufacturing and Renewable Technologies 2021.

conductivity [10,11], and UV stability [12]. Similarly, besides mechanical properties, clay and layered hydroxides can improve the flame retardancy [13] and barrier properties [14,15] of the polymers. The properties exerted by a certain nanofiller are unique for a polymer matrix composite that cannot be replicated by another polymer/nanofiller combination.

EVACO is a semi-crystalline polymer and the presence of HNTs can improve the crystallinity of EVACO as the same is observed in many other semi-crystalline polymer matrices. The presence of carbonyl groups in the backbone of EVACO increases the polarity and thereby its affinity for metallic surfaces [16]. Halloysite is a halogen and phosphorous free flame retardant and the water molecules present between the SiO₄ and AlO₆ layers [17] will dilute the free radicals or the reactive species at the flame front to enter the flame as the combustion begins. Halloysite nanotubes have also found applications, in controlled drug release [18] and protective agents [19], fillers [20–23], emulsifiers [24], adsorbents for pollutants [17], etc.

The characteristics of the polymers depend on their chain dynamic [25] and the motions of the chains are also influenced by nanofillers. The nucleating ability of the nanofillers, at low filler loading, improves the crystallinity of polymer matrices [21,26]. The solution cast EVACO/HNT composites were extensively studied using different characterization techniques. In this work, the effect of the position of HNTs on the crystalline and mechanical properties of EVACO matrix is studied by image processing of elemental mapped electron micrographs. The tensile properties and crystallinity of the composites are affected by the dispersion and distribution of HNTs. The presence of HNTs in the EVACO matrix improved the flame retardancy of the composite.

2. Materials and methods

The materials for the present study were Elvaloy®4924 (EVACO), HNTs, and dichloromethane (DCM), which are purchased from Du Pont India., Sigma Aldrich, India, (product ID: 685445) and Central Drug House (P) Ltd, India, respectively. To prepare the composites, a known quantity of EVACO was dissolved by constant stirring (at a speed of ~700 rpm) into a fixed quantity of DCM using a magnetic stirrer in a closed beaker. A known quantity of HNTs was well dispersed in a small part of DCM by stirring and subsequent ultrasonication for 30 min. The above solution of HNTs in DCM was combined with the former EVACO solution by stirring followed by ultrasound treatment. The mixture was then poured onto glass petri dishes to create the respective composite films and are allowed to dry at room temperature and then in a vacuum oven at 50 °C for 6 h. Composites films with HNT loadings 1, 3, 5, 7, and 10 wt.%, respectively were prepared.

Energy-dispersive X-ray spectroscopy (EDS) (Link ISIS-300, Oxford Instruments, UK) was used to map the aluminum in the composite, each aluminum dot corresponds to HNTs, and Image-J software [27] to analyse the maps. X-ray diffraction patterns (JEOL, DX-GE-2P, Japan) of the composite sheets were analyzed using CuK_α radiation to determine the crystallinity of nanocomposites. The percentage of crystallinity (X_c) was estimated by deconvoluting the XRD patterns to amorphous and crystalline contributions, and the extend of crystallinity was estimated by the ratio [28].

$$X_c = \frac{I_c}{I_a + I_c} \quad (1)$$

where I_a and I_c represent the integrated intensities of the amorphous and crystalline regions in EVACO, respectively.

Fourier transforms infrared (FTIR) spectra (Jasco FTIR 4200, Japan) of the EVACO and the representative composite were

recorded in ATR mode in a wavenumber range of 650–4000 cm⁻¹. Thermogravimetric measurements (TGA Q5000, TA instruments, USA) were performed under nitrogen atmosphere for the samples under a nitrogen flow of 25 mL min⁻¹ and at a constant heating rate of 10 °C min⁻¹. Differential scanning calorimetric measurements (DSC) (Mettler Toledo DSC, USA) were carried out in a nitrogen atmosphere between 0 and 150 °C at a heating rate of 10 °C min⁻¹. The extend of crystallinity of EVACO and the composites were determined from the area under the endothermic curve, using the equation [29]:

$$X_c = \frac{\Delta H_f}{W_i \times \Delta H_{f100\%}} \times 100 \quad (2)$$

where W_i = weight fraction of the polymer, X_c = crystallinity (%), ΔH_f = enthalpy of melting of completely crystalline EVACO (J/g), $\Delta H_{f100\%}$ = enthalpy of crystallization of a 100% crystalline sample of EVA = 68 Jg⁻¹ [30].

The tensile measurements (Hounsfield Universal Testing Machine, H25KS, Hounsfield, UK) at ambient conditions were made for three dumb-bell samples, prepared according to ASTM D 412-B. The fractured surfaces were analyzed using a scanning electron microscope (SEM) (JSM-6380LA, JEOL, Japan). The specimens were sputtered with gold (JEOL JFC 1600) in an auto fine coater before imaging.

3. Results and discussion

3.1. Dispersion and distribution of HNTs in EVACO

To find the dispersion of HNTs, each dot in the aluminum elemental map was considered as an HNT and a sparse sampling technique was employed. The aluminum elemental map of a representative composite is shown in Fig. 1. In sparse sampling, the elemental maps were divided into 20 equal sections and the numbers of particles in each section were counted. The average number of particles per unit area was calculated and the respective standard deviation was estimated for each composite. A large standard deviation shows a poor dispersion and the standard deviation was calculated using the following equation [31]:

$$\sigma = \sqrt{\frac{1}{n} \sum_i (N_{Ai} - \bar{N}_A)^2} \quad (4)$$

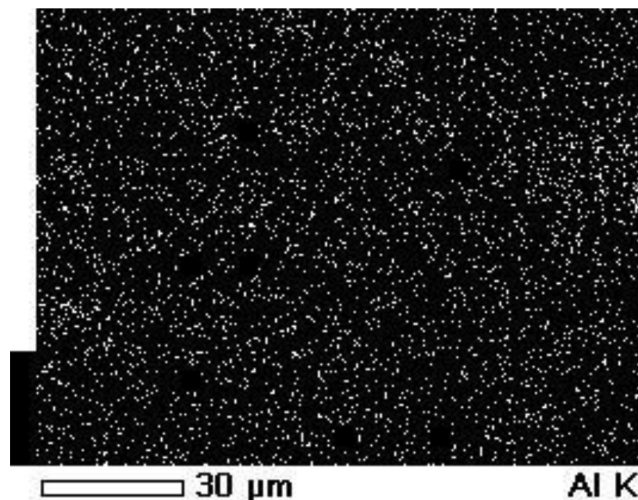


Fig. 1. Representative Aluminium elemental map of the composite with 1 wt.% HNT loading.

where N_{Ai} represents the counts of inclusions per unit area in the i^{th} location, \bar{N}_A is the number of particles in the unit area and σ is the standard deviation.

The sparsely sampled elemental maps of the synthesized composites are shown in Fig. 2. The average number of particles per unit area, standard deviation, and the expected number of particles are presented in Table 1. From Table 1, it is clear that the standard deviation in the number of particles in each section of the composite is increased with filler loading. The increase in standard deviation is due to the agglomeration of the particle that in turn made a significant difference in the average number of particles which was supposed to be increasing akin to the expected number of particles. The expected number of particles was calculated by multiplying the average number of particles in a composite with 1 wt.% filler loading with the higher filler loadings. At 1 wt.% filler loading, HNTs were dispersed uniformly.

To understand the distribution of the particle, the distance between each particle and its nearest neighbor (NND) was calculated using an ImageJ plug-in [32] from the aluminum elemental maps. Fig. 3 shows the distribution of nearest neighbor distances for a representative composite. The nearest neighbor distance compares the position of a particular nanotube with respect to the other nanotubes in the composite. For uniform distribution, the nearest neighbor distribution should be narrow and it was estimated by calculating the FWHM of the Gaussian fit of the distribution and it is presented in Table 2.

The ratio of the average actual neighbor distance (R_k) to the average expected nearest neighbor distance (E_k) [33] of the particles in the composites is another means of optimizing the NND. Higher the R_k/E_k ratio, the better the distribution. R_k and E_k are estimated using the following equation.

$$R_k = \frac{\sum_i^n d_i}{n} \text{ and } E_k = \frac{0.5}{\sqrt{\frac{n}{A}}} \quad (5)$$

where A is the area under study, d_i is the distance between the i^{th} particle and its imminent particle, and n is the number of particles.

In EVACO/HNT composites, the NND of 1 wt.% HNT loading has a wider distribution than that of a 3 wt.% HNT loaded composite. Since 1 wt.% HNT loaded composite has less number of HNTs in it. For filler loadings above 3 wt.%, NND has a broad distribution, which is due to the presence of agglomerates. R_k/E_k ratio is also

Table 1
Sparse sampling.

Filler loading (wt.%)	The average number of particles unit area	Standard deviation in the number of particles	Expected number of particles.
1	85	8	85
3	98	9	255
5	99	14	425
7	89	18	595
10	80	24	850

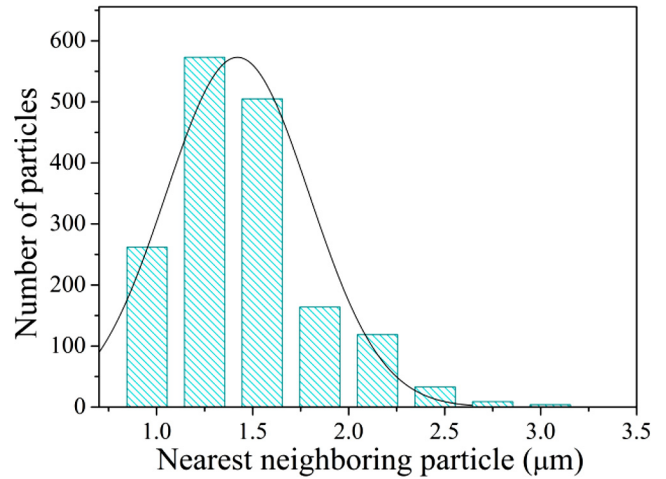


Fig. 3. Nearest neighbor distribution of 1 wt.% HNT filled nanocomposite.

Table 2
Nearest neighbor distance.

HNT loading (wt.%)	FWHM of NND	R_k/E_k
1	0.88	1.365
3	0.74	1.344
5	0.79	1.331
7	0.88	1.292
10	0.94	1.237

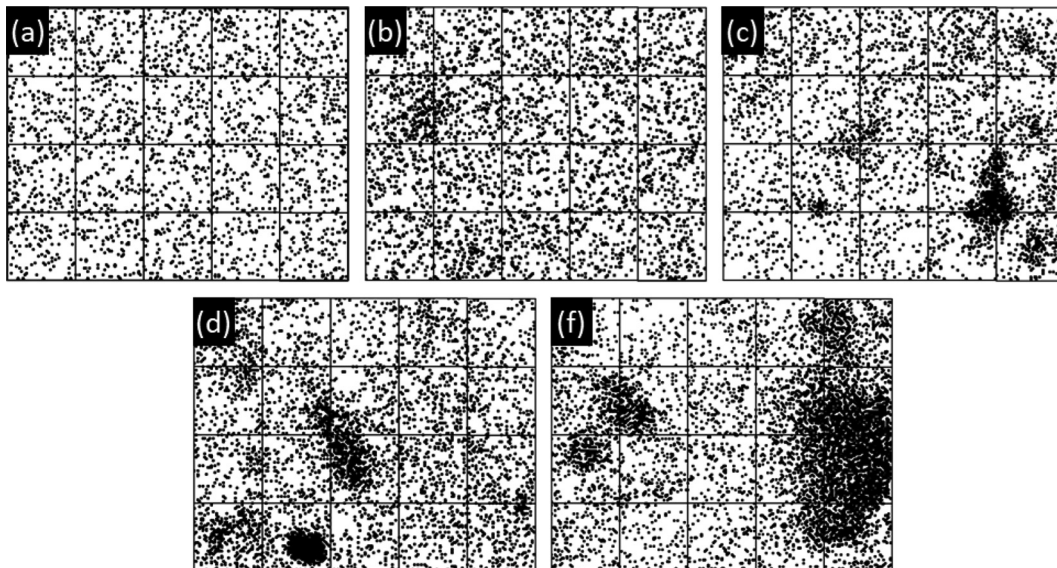


Fig. 2. Sparse sampled aluminum maps of HNT loaded composites, (a) 1 wt.%, (b) 3 wt.% (C), 5 wt.% (d) 7 wt.% and (e) 10 wt.%.

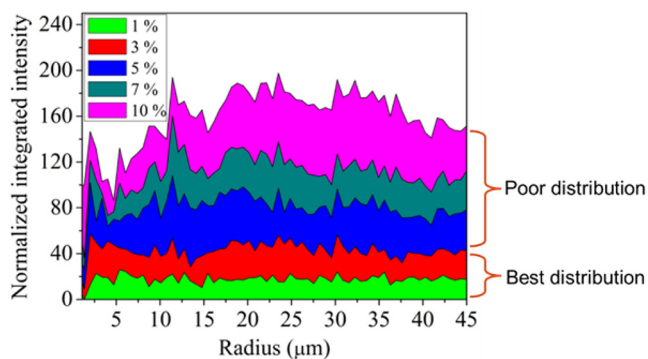


Fig. 4. Radial distribution of HNTs in EVACO/HNT composites.

decreasing with the filler loading since the agglomerates can impact the uniform dispersion and distribution of the fillers in the polymer matrix.

The radial distribution function also specifies the distribution of the nanotubes in the composites. It is a measure of the number density of the particles along the radial direction with respect to a reference particle [34].

The uniformity in the radial distribution of HNTs is reduced as the weight percent of the nanofiller is increased (Fig. 4). The straight radial map shows a uniform distribution, whereas the rough map stands for the non-uniform distribution of HNTs. The indistinguishable boundary of the HNTs in the TEM image (Fig. 5) of the selected composite reveals the interaction between the filler and the matrix. The crystallization of the polymer around HNTs tells the nucleating ability of the HNTs. The major vibrational peaks corresponding to the functional groups of pristine EVACO is obtained through FTIR analysis and it is presented in Table 3.

3.2. The crystallinity of EVACO/HNT composite

The influence of HNTs on the crystallinity of EVACO was evaluated by DSC and XRD analysis. The percentage crystallinity obtained from XRD and DSC analysis is presented in Table 4. The broad melting peak between 20 and 100 °C in Fig. 6 is due to the uneven random crystals in the semi-crystalline EVACO. The addition of small quantities of HNTs i.e., 1 wt.% and 3 wt.% improves

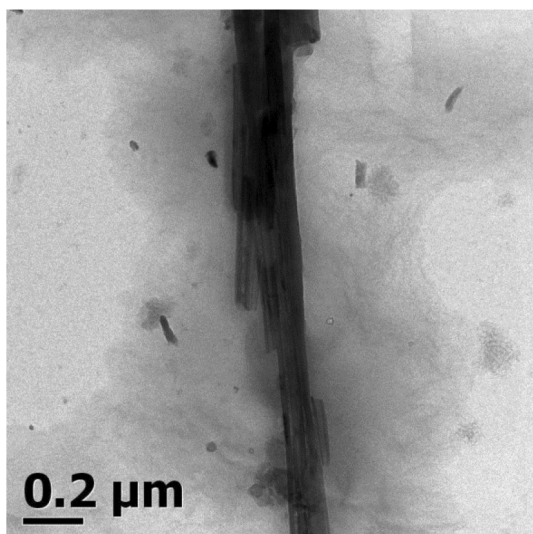


Fig. 5. TEM micrograph of EVACO/CNT composite with 1 wt.% HNT loading.

Table 3
FTIR spectra peak positions.

Peak position (cm ⁻¹)	Assignment
3418	—OH stretching
2919 and 2850	Symmetric and asymmetric CH ₂ stretching
1736	C=O stretching
1467 and 1375	CH scissoring and symmetric deformation
1231	Twisting and wagging of CH
1019	C—OH stretching
721	Rocking vibration of CH

Table 4
Percentage crystallinity of EVACO and EVACO/HNT composites.

Filler loading (wt.%)	% crystallinity from DSC	% crystallinity from XRD
0	46.84	26.59
1	50.33	28.96
3	51.77	30.21
5	46.44	25.75
7	42.13	24.71
10	41.88	18.71

the crystallinity in the composite remarkably. The uniform dispersion and distribution of HNTs can be attributed to this increment. If the filling of the filler is increased above 3 wt.%, the nanotubes may arrest the spherulitic growth front, which originates from the nucleation source, thereby reducing the growth of crystalline regions and ultimately a reduction in the crystallinity. Additionally, a large number of tubes in the matrix can decrease the movement of polymer chains, which can otherwise undergo crystallization in the absence of halloysite nanotubes. The large agglomerates can also adversely affect the crystallinity of the composite with high halloysite nanotube loading.

The XRD results also show the increase in crystallinity for filler loadings of 1 wt.% and 3 wt.% and decrease thereafter. The discrepancy in percentage crystallinity from the study of DSC and XRD is due to the eventual errors that can arise during XRD pattern deconvolution and baseline line correction in DSC curves. [35]. It can be concluded that the dipole–dipole attraction between the nanofiller and the matrix at low nanofiller loading, especially when they are in the solution, can bring the polymer chains close together and align them in an order to favor crystallization.

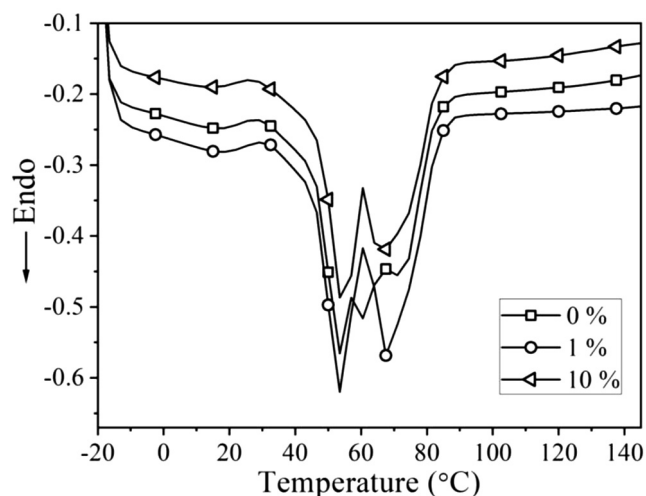


Fig. 6. DSC first heating curves of pristine EVACO and representative composites.

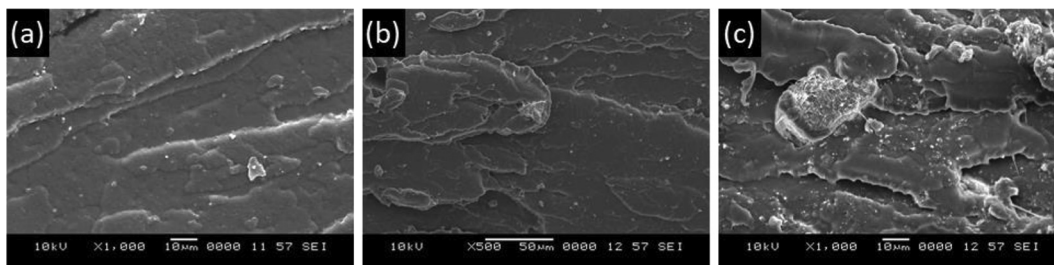


Fig. 7. The fracture surface of (a) EVACO, (b) EVACO/HNT composite with 1 wt.% HNT loading, and (c) 7 wt.% HNT loading.

Table 5

Tensile properties of virgin EVACO and its nanocomposites.

Filler loading (wt.%)	Tensile strength (MPa)	Toughness (kJ/m)
0	31.5	14.3
1	33.2	16.2
3	31.2	14.1
5	29.2	13.5
7	27.2	12.2
10	25.9	8.3

3.3. Mechanical properties

The pristine and composite samples of EVACO exhibited ductile fracture, as it can be identified by the continuous crack propagation trajectories on the fractured surfaces (Fig. 7). Due to the good dispersion and distribution of halloysite nanotubes, the highest tensile strength is observed for the nanocomposite with 1 wt.% percent filler loading (Table 5). In the composite with 3 wt.% halloysite nanotube loading, an overall increase in crystallinity is found, but the dispersion as well as the distribution of the halloysite nanotubes is inferior to 1 wt.% HNT loading. The resulting non-uniform distribution of stress ultimately leads to a small decrease in the ultimate tensile strength. At high filler loading, >3 wt.%, the cluster of halloysite nanotubes and the debonding of these agglomerations from the polymer leads to premature failure, as observed in SEM micrographs of the fracture surface in Fig. 7, and the lessening in the ultimate tensile strength.

4. Conclusions

It is summarised that the dispersion and distribution of the filler play a key role in controlling the crystallizability and mechanical characteristics of the Elvaloy[®]4924 (EVACO)/halloysite nanotube nanocomposites. The image processing of SEM-elemental maps revealed that 1 wt.% HNT loading shows a good dispersion and distribution of the fillers in the matrix. The reduction in mechanical and crystalline characteristics of the composites are in good agreement with dispersion and the uniform spreading of halloysite nanotubes in an array. For 1 wt.% HNT loading, the composite exhibits the best mechanical characteristics and crystallinity. The halloysite nanotubes influence the crystallinity of EVACO at low filler weight fractions, thus discloses the halloysite nanotube's potential as a nucleating agent.

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.

Acknowledgments

The authors greatly appreciate the continuous support from the Departments of Mechanical Engineering and Automobile Engineering, SCMS School of Engineering and Technology, Ernakulam, India, and the management of SCMS Group of Educational Institutions, Ernakulam India.

References

- [1] P. Chaudhary, F. Fatima, A. Kumar, Relevance of nanomaterials in food packaging and its advanced future prospects, *J. Inorg. Organomet. Polym. Mater.* 30 (12) (2020) 5180–5192.
- [2] S. Fu, Z. Sun, P. Huang, Y. Li, N. Hu, Some basic aspects of polymer nanocomposites: a critical review, *Nano Mater. Sci.* 1 (1) (2019) 2–30.
- [3] V.K. Sharma, J. Filip, R. Zboril, R.S. Varma, Natural inorganic nanoparticles – formation, fate, and toxicity in the environment, *Chem. Soc. Rev.* 44 (23) (2015) 8410–8423.
- [4] J.L. Suter, D. Groen, P.V. Coveney, Mechanism of exfoliation and prediction of materials properties of clay-polymer nanocomposites from multiscale modeling, *Nano Lett.* 15 (12) (2015) 8108–8113.
- [5] D.R. Paul, L.M. Robeson, *Polymer nanotechnology: nanocomposites*, *Polymer* 49 (2008) 3187–3204.
- [6] D.R. Baer, M.H. Engelhard, G.E. Johnson, J. Laskin, J. Lai, K. Mueller, P. Munusamy, S. Thevuthasan, H. Wang, N. Washton, A. Elder, B.L. Baisch, A. Karakoti, S.V.N.T. Kuchibhatla, D. Moon, Surface characterization of nanomaterials and nanoparticles: Important needs and challenging opportunities, *J. Vac. Sci. Technol. Vac. Surf. Films Off. J. Am. Vac. Soc.* 31 (5) (2013) 050820, <https://doi.org/10.1116/1.4818423>.
- [7] I.M. Hamouda, Current perspectives of nanoparticles in medical and dental biomaterials, *J. Biomed. Res.* 26 (2012) 143–151.
- [8] R. Andrews, M.C. Weisenberger, Carbon nanotube polymer composites, *Curr. Opin. Solid State Mater. Sci.* 8 (1) (2004) 31–37.
- [9] J.N. Coleman, U. Khan, W.J. Blau, and Y.K. Gun'ko: Small but strong: A review of the mechanical properties of carbon nanotube-polymer composites. *Carbon* 44(9), 1624–1652 (2006).
- [10] C. Min, X. Shen, Z. Shi, L. Chen, Z. Xu, The electrical properties and conducting mechanisms of carbon nanotube/polymer nanocomposites: a review, *Polym.-Plast. Technol. Eng.* 49 (12) (2010) 1172–1181.
- [11] N. Hu, Z. Masuda, C. Yan, G. Yamamoto, H. Fukunaga, T. Hashida, The electrical properties of polymer nanocomposites with carbon nanotube fillers, *Nanotechnology*. 19 (2008) 215701.
- [12] S. Morlat-Therias, E. Fanton, J.-L. Gardette, S. Peeterbroeck, M. Alexandre, P. Dubois, Polymer/carbon nanotube nanocomposites: Influence of carbon nanotubes on EVA photodegradation, *Polym. Degrad. Stab.* 92 (10) (2007) 1873–1882.
- [13] A.B. Morgan, C.A. Wilkie, *Flame Retardant Polymer Nanocomposites*, Wiley-Blackwell, New Jersey, 2007.
- [14] D. Feldman, Polymer nanocomposite barriers, *J. Macromol. Sci. Part A*. 50 (4) (2013) 441–448.
- [15] V. Mittal, *Barrier properties of Polymer Clay Nanocomposites*, Nova Science Pub. Inc., New York, 2010.
- [16] J.O. Emslander: Image receptor medium containing ethylene vinyl acetate carbon monoxide terpolymer. U.S. Patent, WO2000052532 A1 (2000).
- [17] M. Zhao, P. Liu, Adsorption behavior of methylene blue on halloysite nanotubes, *Microporous Mesoporous Mater.* 112 (1-3) (2008) 419–424.
- [18] N.G. Veerabadran, R.R. Price, Y.M. Lvov, Clay nanotubes for encapsulation and sustained release of drugs, *Nano* 02 (02) (2007) 115–120.
- [19] Y.M. Lvov, D.G. Shchukin, H. Möhwald, R.R. Price, Halloysite clay nanotubes for controlled release of protective agents, *ACS Nano* 2 (5) (2008) 814–820.
- [20] H. Ismail, S.Z. Salleh, Z. Ahmad, Properties of halloysite nanotubes-filled natural rubber prepared using different mixing methods, *Mater. Des.* 50 (2013) 790–797.
- [21] M. Liu, B. Guo, M. Du, F. Chen, D. Jia, Halloysite nanotubes as a novel β -nucleating agent for isotactic polypropylene, *Polymer* 50 (13) (2009) 3022–3030.

- [22] E. Abdullayev, V. Abbasov, A. Tursunbayeva, V. Portnov, H. Ibrahimov, G. Mukhtarova, Y. Lvov, Self-healing coatings based on halloysite clay polymer composites for protection of copper alloys, *ACS Appl. Mater. Interfaces*. 5 (10) (2013) 4464–4471.
- [23] L.A. Dobrzański, B. Tomiczek, M. Adamiak, K. Golombek, Mechanically milled aluminium matrix composites reinforced with halloysite nanotubes, *J. Achiev. Mater. Manuf. Eng.* 55 (2012) 7.
- [24] Z. Wei, C. Wang, H. Liu, S. Zou, Z. Tong, Halloysite nanotubes as particulate emulsifier: preparation of biocompatible drug-carrying PLGA microspheres based on pickering emulsion, *J. Appl. Polym. Sci.* 125 (2012) E358–E368.
- [25] M.F. Talbot, G.S. Springer, L.A. Berglund, The effects of crystallinity on the mechanical properties of PEEK polymer and graphite fiber reinforced PEEK, *J. Compos. Mater.* 21 (11) (1987) 1056–1081.
- [26] E. Assouline, A. Lustiger, A.H. Barber, C.A. Cooper, E. Klein, E. Wachtel, H.D. Wagner, Nucleation ability of multiwall carbon nanotubes in polypropylene composites, *J. Polym. Sci. Part B Polym. Phys.* 41 (2003) 520–527.
- [27] J.M.M. Perez, J. Pascau, *Image Processing with ImageJ*, Packt Publishing, 2013.
- [28] S. Park, J.O. Baker, M.E. Himmel, P.A. Parilla, D.K. Johnson, Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance, *Biotechnol. Biofuels* 3 (2010) 10.
- [29] Y. Kong, J.N. Hay, The measurement of the crystallinity of polymers by DSC, *Polymer* 43 (14) (2002) 3873–3878.
- [30] S. Chattopadhyay, T.K. Chaki, A.K. Bhowmick, Heat shrinkability of electron-beam-modified thermoplastic elastomeric films from blends of ethylene-vinyl acetate copolymer and polyethylene, *Radiat. Phys. Chem.* 59 (5-6) (2000) 501–510.
- [31] J.J. Friel, A.S.M. International, *Practical Guide to Image Analysis*, ASM International, 2000.
- [32] Y. Mao, Nearest Neighbor Distances Calculation with ImageJ - EVOCD, (2016). https://icme.hpc.msstate.edu/mediawiki/index.php/Nearest_Neighbor_Distances_Calculation_with_ImageJ.html (accessed February 26, 2021).
- [33] P.J. Clark, F.C. Evans, Distance to nearest neighbor as a measure of spatial relationships in populations, *Ecology* 35 (1954) 445–453.
- [34] P. Baggethun, Radial Profile Plot, (2002). https://imagejdocu.tudor.lu/macro/radial_distribution_function (accessed February 26, 2021).
- [35] M.F.S. Lima, M.A.Z. Vasconcellos, D. Samios, Crystallinity changes in plastically deformed isotactic polypropylene evaluated by x-ray diffraction and differential scanning calorimetry methods, *J. Polym. Sci. Part B Polym. Phys.* 40 (9) (2002) 896–903.



Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Effect of nanofillers on the crystalline and mechanical properties of EVACO polymer nanocomposites

Gibin George^{a,*}, H. Manikandan^a, T.M. Anup Kumar^a, Sam Joshy^a, A.C. Sanju^a, S. Anandhan^{b,*}

^aDept. of Mechanical Engineering, SCMS School of Engineering and Technology, Pallissery, Ernakulam, Kerala, India

^bDept. of Metallurgical and Materials Engg, National Institute of Technology Karnataka, Surathkal, Karnataka, India

ARTICLE INFO

Article history:

Received 3 March 2021

Received in revised form 12 April 2021

Accepted 15 April 2021

Available online xxx

Keywords:

Nanocomposite

Crystallization

Nanofillers

Polymer composite

ABSTRACT

In this work, the effect of different fillers on the crystalline and mechanical properties of the poly (ethylene-co-vinyl acetate-co-carbon monoxide) (EVACO) terpolymer composite is studied systematically. Alumina trihydrate nanoparticles (nano-ATH), halloysite nanotubes (HNTs), and the multiwalled carbon nanotubes (MWCNTs) are the representative fillers used in the present study. The surface of MWCNTs are decorated using carbonyl, however, nano-ATH and HNTs are used without any surface treatment. The mechanical properties of the composites are evaluated using a tensile test and the improvement in the mechanical properties can be correlated to the improvement in the crystallinity in the composite. The presence of nanofillers in the EVACO matrix significantly influenced the crystallinity, which was determined by X-ray diffraction. The fractography studies reveal the presence of agglomerates at high filler loading results in the subsequent reduction in the tensile properties. Interestingly, the MWCNTs at very low filler loading significantly enhances the tensile properties of EVACO.

© 2021 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the International Conference on Sustainable materials, Manufacturing and Renewable Technologies 2021.

1. Introduction

Polymer nanocomposites are finding new applications every day and replacing the conventional polymers from household to advanced engineering applications [1]. Nanofillers from various sources/origin are commonly used as fillers in polymer nanocomposites. The nanofillers with superior mechanical properties are often used in polymers with poor mechanical characteristics. The addition of nanomaterials in the polymer matrix can impart certain unique properties that cannot be matched by any other material. In the case of nanocomposites, the properties of polymers and nanofillers are often compromised and they exhibit superior properties as a combined material [2]. Additionally, a small quantity of the nanofiller is sufficient to make a significant impact on the properties of the polymer matrix.

Nanomaterials, such as carbon nanotubes [3], clay [4], alumina trihydrate [5], layered double hydroxides [6], halloysite nanotubes

[7], nanocellulose [8], graphene [9], etc. are the common multifunctional nanofillers used in the polymer nanocomposites. The properties such as crystallinity, thermal degradation, tensile strength, permeation resistance, electrical conductivity, flame retardance, etc. are affected by the addition of nanofillers. The unique characteristics of the fillers impart significant property enhancement in the polymer nanocomposite without affecting the processability of the polymer. For instance, the two-dimensional layered nanostructures can influence the permeation characteristics of the polymer [10]. Similarly, the carbon nanotubes increase the electrical conductivity [11], and alumina trihydrate imparts flame retardancy [12]. The aspect ratio (AR) of the nanofillers also impacts the mechanical properties of the polymer nanocomposites [13]. The proven ability of nanofillers as nucleating agents to improve the crystallinity of several semi-crystalline polymer matrices [14–16] that in turn contribute to the enhancement in the tensile strength of polymer composites.

Interfacial bonding between the matrix and the nanofiller plays an important role in determining the properties of the nanocomposites. The interfacial bonding of the filler and the matrix can be improved by modifying the polymer or the filler with suitable functional groups. However, modifying the filler is easier than

* Corresponding authors.

E-mail addresses: gibingeorge@scmsgroup.org (G. George), manikandan@scmsgroup.org (H. Manikandan), anupkumartm@scmsgroup.org (T.M.A. Kumar), samjoshy@scmsgroup.org (S. Joshy), sanju@scmsgroup.org (A.C. Sanju), anandtm@gmail.com, anandhan@nitk.edu.in (S. Anandhan).

<https://doi.org/10.1016/j.matpr.2021.04.613>

2214-7853/© 2021 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the International Conference on Sustainable materials, Manufacturing and Renewable Technologies 2021.

the modification of polymer, and the fillers are often modified to match the polarity of the polymer matrix.

A polar filler is modified with a non-polar agent to be used in a non-polar polymer matrix, but it can be directly used in a polar polymer. Additionally, the dispersion of the nanofillers also impacts the mechanical properties of the nanocomposites [17]. Surface modification of carbon nanotubes (MWCNTs) is essential before it is mixed with the organic matrices, since pristine MWCNTs exist as bundles due to their inertness [18]. The polar nanofillers such as ATH and HNTs can be directly used as nanofillers into a polar polymer matrix.

A carbonyl group is introduced to the copolymer poly(ethylene-co-vinyl acetate) (EVA) to form the terpolymer poly(ethylene-co-vinyl acetate-co-carbon monoxide) (EVACO). Such an addition increases the polarity of the new polymer. The polarity of EVA is difficult to improve by increasing the vinyl acetate content alone, as the increase in vinyl acetate content can adversely affect the properties of EVA [19]. The addition of carbon monoxide to the backbone of EVA increases the polarity of the polymer, thereby improves its adhesion to polar surfaces, therefore EVACO is used as an adhesion booster in coatings. EVACO is semi-crystalline in nature, and the polyethylene phase imparts crystallization in it.

In this study, EVACO/nanofiller composites are prepared through solution casting. Industrial processing of EVACO is mainly in the form of solutions, and the method used here is akin to the bulk processing of EVACO. The polar fillers such as ATH and HNTs are directly reinforced to the EVACO matrix, but, MWCNTs are surface modified with polar functional groups before reinforcing them into the EVACO matrix. The mechanical properties and crystallizability of EVACO can be improved by the addition of nanofillers in small quantities. The effect of different nanofillers on the mechanical properties and crystallinity of EVACO is studied here.

2. Materials and methods

Poly(ethylene-co-vinyl acetate-co-carbon monoxide) (Elvaloy® 4924) provided by Du Pont, India, halloysite nanotubes (HNTs) and carbon nanotubes (MWCNTs) procured from Sigma Aldrich, India, nano-ATH obtained from US Research Nanomaterials Inc., USA, and dichloromethane (DCM) procured from Central Drug House (P) Ltd, New Delhi, India were used in the present study.

To fabricate the composites, a predetermined quantity of EVACO is dissolved in DCM by continuous mixing using a magnetic stirrer. To the above solution, the appropriate quantities of nanofillers are slowly added and mixed thoroughly by vigorous stirring and subsequent ultrasonication. The mixture is then poured into Petri-dishes and allowed to dry to get the respective composite films. The composite films are then dried at room temperature and then in a vacuum oven at 50 °C for 6 h.

The tensile measurements (Hounsfield Universal Testing Machine, H25KS, Hounsfield, UK) at ambient conditions were made for three dumb-bell samples, at a crosshead speed of 50 mm/minute, prepared according to ASTM D 412-B. The fractured surfaces were analyzed using a scanning electron microscope (SEM) (JSM-6380LA, JEOL, Japan) and the samples were sputtered with gold (JEOL JFC 1600), in an auto fine coater, prior to imaging. Transmission electron microscope, CM12 PHILIPS, Netherlands, was used to image the morphology of the nanofillers. X-ray diffraction patterns (JEOL, DX-GE-2P, Japan) of the composite sheets were analyzed using CuK_α radiation to determine the crystallinity of nanocomposites. The degree of crystallinity (X_c) was calculated by deconvoluting the XRD patterns to amorphous and crystalline contributions and the degree of crystallinity was calculated by the ratio [20].

$$X_c = \frac{I_c}{I_a + I_c} \quad (1)$$

where I_a and I_c are the integrated intensities corresponding to the amorphous and crystalline phases, respectively.

3. Results and discussion

3.1. Morphology of fillers

The morphology of the nanofillers is compared in Fig. 1 a–c, and one can observe a significant difference in the aspect ratio (AR) of the nanofillers used in the present study. On comparing the TEM images of the nanofillers, the highest AR is observed in the case of MWCNTs (AR≈100) (Fig. 1a), followed by HNTs (AR≈20) (Fig. 1b), and the lowest in the case of nano-ATH (AR≈1.0) (Fig. 1c). Additionally, the surface texture of the MWCNTs is smoother than HNTs and ATH besides the smaller diameter.

3.2. Crystallinity of the nanocomposites

The crystallinity of the semicrystalline polymers can be influenced by the nanofillers. Many nanofillers act as nucleating agents when they are incorporated into different semicrystalline polymer matrices [21,22]. Due to the high polar nature of the nanofillers used in the present study, the crystallizable polymer chains are pulled together to form for more crystallites, that are not formed in the absence of nanofillers [23,24]. The percentage crystallinity is increasing with the filler loading initially but decreases thereafter.

From Table 1, in the case of each nanofiller type, as the filler loading is increased, the corresponding crystallinity (X_c) was initially increased (at 1% loading of ATH and HNTs, and 0.1 wt% loadings in the case of MWCNTs). The crystallinity is decreasing slowly after the above threshold of filler loading in the case of all the fillers. It is likely that, at high filler loading, due to agglomerations, the nanoparticles hinder the polymer chain movements and reduces the crystallinity. One can note that, in the case of MWCNT loading, even though the wt% of MWCNT loading is far less than HNTs or nano-ATH loading, the change in crystallinity is analogous to the other experiments. Apparently, the crystallinity of the present nanocomposites is a function of filler loading besides the nature of the fillers.

3.3. Mechanical properties

In general, all the semicrystalline polymers exhibit a ductile fracture. EVACO is a semicrystalline polymer, therefore, large plastic deformation is expected for pristine EVACO and its nanocomposites. The ductility of the composite decreases as the filler loading exceeds a certain threshold. In general, irrespective of the wt% of the nanofiller, the mechanical properties decrease after an initial surge for all the nanocomposite variants. The maximum value of ultimate tensile strength and toughness is observed for 1% filler loading in both HNT and ATH nanocomposites, whereas for MWCNT reinforced nanocomposites the the highest tensile properties are observed for 0.05% filler loading, as shown in Table 2. The reduction in the mechanical properties with an increase in filler loadings accounts for the agglomeration of nanoparticles leading to non-uniform stress distribution in the composites. The low filler loading of MWCNTs makes a significant impact on the tensile properties of EVACO than HNTs or nano-ATH, since MWCNTs have a significantly higher aspect ratio, as shown in Fig. 1a, and lower density ($\rho = 2.1$ g/cc) as compared to HNTs ($\rho = 2.53$ g/cc) and ATH ($\rho = 2.42$ g/cc). Moreover, the tensile strength of individual strands of MWCNTs is equivalent to that of a steel wire with the same dimensions. Additionally, the dispersion of the nanofillers is uniform at low filler loading, consequently, the effect of stress concentration by agglomeration and the associated premature

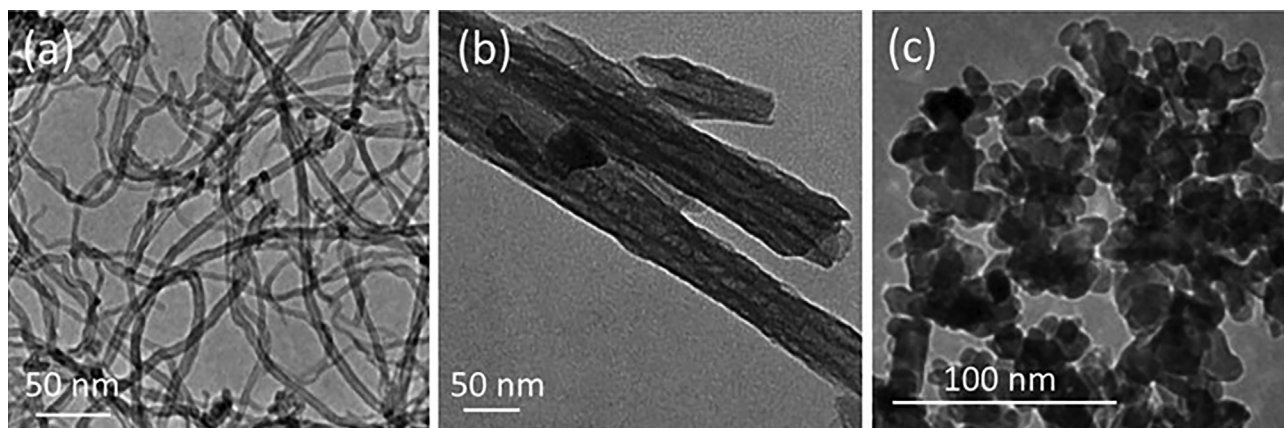


Fig. 1. TEM images of (a) MWCNTs, (b) HNTs and (c) nano-ATH.

Table 1
Normalized percentage of crystallinity for each filler loading.

Filler	Sl. No	Filler Loading (wt%)	X_c
ATH	1	0	46.84
	2	1	50.33
	3	3	51.77
	4	5	46.44
	5	7	42.13
HNTs	1	0	46.84
	2	1	56.98
	3	3	53.69
	4	5	42.73
	5	7	37.69
MWCNTs	1	0	46.84
	2	0.05	48.76
	3	0.1	52.51
	4	0.15	49.67
	5	0.2	48.31
	6	0.25	45.25

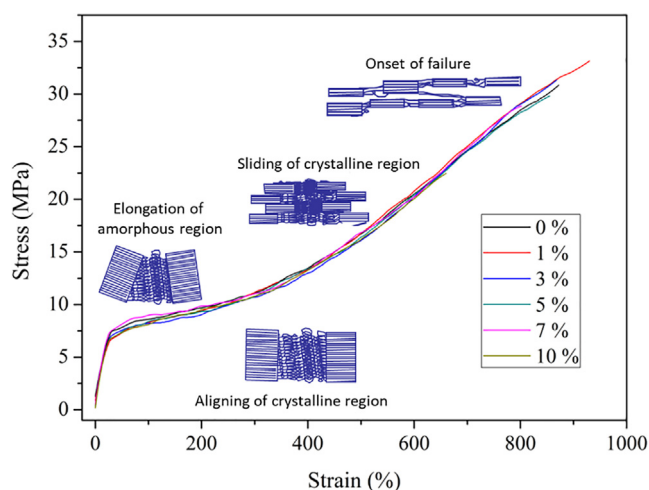


Fig. 2. Representative stress-strain curve of EVACO/nanofiller composite.

failure can be mitigated in those composites. In all the cases, the tensile strength increases initially and decreases thereafter as the filler loading is increased. The representative stress-strain curve of EVACO/HNTs nanocomposites is shown in Fig. 2. There is no apparent change in the profile with filler loading, however, the strain is improved after the filler addition, as compared with the pristine polymer. The schematic diagram in the inset shows the behavior of crystallites in a semi-crystalline polymer as the applied

stress increases. Which in turn indicate the role of crystallinity on the tensile properties of the polymer nanocomposites.

3.4. Fractography analysis

The fracture surfaces of the representative nanocomposites after the tensile test is observed using SEM, as shown in Fig. 3a-c. From

Table 2
Comparison of mechanical properties of EVACO/nanofiller composites.

Type of filler	Filler loading (wt%)	Ultimate tensile strength (MPa)	Toughness (kN m^{-1})	Percentage elongation at break (%)
Nano-ATH	0	31.5	14.3	834
	1	33.2	16.2	893
	3	31.2	14.1	821
	5	29.2	13.5	748
	7	27.2	12.2	711
	10	25.9	8.30	834
HNTs	0	31.5	14.3	930
	1	34.2	16.8	882
	3	32.0	14.5	845
	5	28.5	12.4	802
	7	26.5	9.9	834
	0	31.5	14.3	944
MWCNTs	0.05	45.6	24.3	924
	0.1	41.2	20.9	921
	0.15	40.5	19.8	913
	0.2	39.9	18.6	902
	0.25	39.4	18.0	910

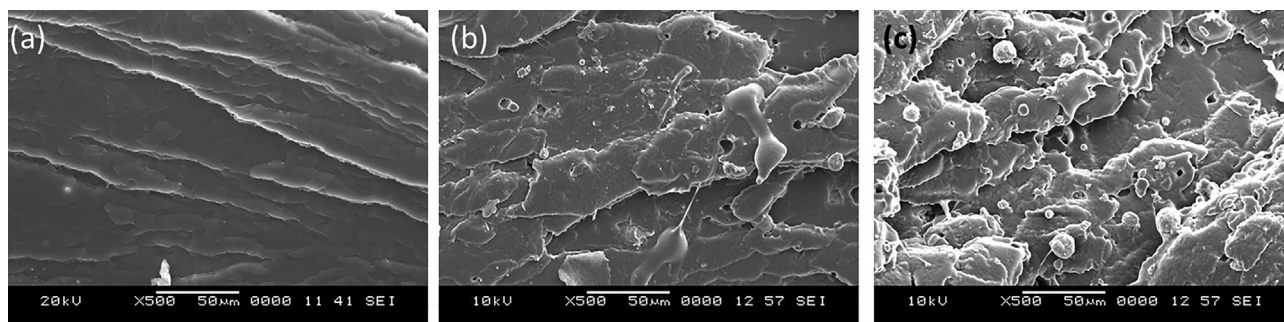


Fig. 3. The fracture surface of (a) EVACO, (b) EVACO/1 wt% ATH composite, and (c) EVACO/10 wt% ATH composite.

the above figures, one can conclude that the fillers significantly affect the mechanism of fracture. In the case of pristine EVACO, the crack propagation trajectories are very clearly indicating a pure ductile failure. As the filler loading is increased to 1 wt% stress whitened regions are observed, which is an indication of more crystallite deformation due to high crystallinity of the composite at 1 wt% loading. In general, the stress whitened regions appears as white regions on the fractured surface of a semicrystalline polymer, which are formed due to the elongation of polymeric chains forming the crystalline part of the polymer. At 10 wt% loading filler loadings, the agglomerated particles and the debonding of these agglomerates from the matrix is clearly visible. The formation of these agglomerates leads to premature failure, resulting in a lower tensile strength than the pristine polymer as observed previously. It is important to note that the roughness of the fractured surfaces are increasing with the filler loading, which conveys the brittle nature of the composites at high filler loading.

4. Conclusions

In summary, the mechanical properties of the EVACO nanocomposites are dependent on the nature of fillers and the wt.% of filler loading. In the case of a semicrystalline polymer, the ability of the nanofillers to form crystallites determines the overall mechanical properties. Thus the filler aspect ratio and the properties of the fillers, in turn, affect the optimal filler loading and the tensile properties. The good dispersion and distribution of fillers also play a major role in controlling the crystallizability and ultimately the mechanical properties. In this study, 1 wt% loading of both HNTs and nano-ATH results in the highest tensile properties in the case of EVACO/HNTs and EVACO/nano-ATH nanocomposites and 0.05 wt% MWCNTs in the case of EVACO/MWCNTs nanocomposites. The improvement in the mechanical properties of EVACO/MWCNTs is way higher at a very low MWCNT loading as compared to HNT and nano-ATH loaded composites. It is assumed that the high aspect ratio and the mechanical properties of MWCNT result in a significant improvement in the tensile properties.

CRedit Author statement

Gibin George: Conceptualization, Data collection, Formal analysis and manuscript drafting. **H. Manikandan:** Manuscript correction, structuring and reviewing. **T.M. Anup Kumar:** Article preparation and data analysis. **Sam Joshy:** Reviewing and editing of manuscript. **A.C.Sanju:** Date interpretation and revision of the manuscript. **S. Anandhan** Designed the experiment, data analysis and article revision.

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.

Acknowledgements

The authors greatly appreciate the continuous support from the Department of Mechanical Engineering, SCMS School of Engineering and Technology, Ernakulam, India and the management of SCMS Group of Educational Institutions, Ernakulam India.

References

- [1] B. Ates, S. Koytepe, A. Ulu, C. Gurses, V.K. Thakur, Chemistry, structures, and advanced applications of nanocomposites from biorenewable resources, *Chem. Rev.* 120 (17) (2020) 9304–9362.
- [2] F. Hussain, M. Hojjati, M. Okamoto, R.E. Gorga, Review article: polymer–matrix nanocomposites, processing, manufacturing, and application: an overview, *J. Compos. Mater.* 40 (2006) 1511–1575.
- [3] C.A. Hewitt, A.B. Kaiser, S. Roth, M. Craps, R. Czerw, D.L. Carroll, Multilayered carbon nanotube/polymer composite based thermoelectric fabrics, *Nano Lett.* 12 (3) (2012) 1307–1310.
- [4] F. Gao, Clay/polymer composites: the story, *Mater. Today.* 7 (11) (2004) 50–55.
- [5] X. Zhang, F. Guo, J. Chen, G. Wang, H. Liu, Investigation of interfacial modification for flame retardant ethylene vinyl acetate copolymer/alumina trihydrate nanocomposites, *Polym. Degrad. Stab.* 87 (3) (2005) 411–418.
- [6] H. Pang, Y. Wu, X. Wang, B. Hu, X. Wang, Recent advances in composites of graphene and layered double hydroxides for water remediation: a review, *Chem. – Asian J.* 14 (15) (2019) 2542–2552.
- [7] M. Liu, Z. Jia, D. Jia, C. Zhou, Recent advance in research on halloysite nanotubes–polymer nanocomposite, *Prog. Polym. Sci.* 39 (8) (2014) 1498–1525.
- [8] H.-M. Ng, L.T. Sin, S.-T. Bee, T.-T. Tee, A.R. Rahmat, Review of nanocellulose polymer composite characteristics and challenges, *Polym.-Plast. Technol. Eng.* 56 (7) (2017) 687–731.
- [9] X. Ji, Y. Xu, W. Zhang, L. Cui, J. Liu, Review of functionalization, structure and properties of graphene/polymer composite fibers, *Compos. Part Appl. Sci. Manuf.* 87 (2016) 29–45.
- [10] G. Choudalakis, A.D. Gotsis, Permeability of polymer/clay nanocomposites: a review, *Eur. Polym. J.* 45 (4) (2009) 967–984.
- [11] Y. Zeng, P. Liu, J. Du, L. Zhao, P.M. Ajayan, H.-M. Cheng, Increasing the electrical conductivity of carbon nanotube/polymer composites by using weak nanotube–polymer interactions, *Carbon* 48 (12) (2010) 3551–3558.
- [12] P.V. Bonsignore, Flame retardant flexible polyurethane foam by post-treatment with alumina trihydrate/ latex binder dispersion systems, *J. Cell. Plast.* 15 (3) (1979) 163–179.
- [13] G.S. Ananthapadmanabha, V. Deshpande, Influence of aspect ratio of fillers on the properties of acrylonitrile butadiene styrene composites, *J. Appl. Polym. Sci.* 135 (2018) 46023.
- [14] J.E.K. Schawe, P. Pötschke, I. Alig, Nucleation efficiency of fillers in polymer crystallization studied by fast scanning calorimetry: carbon nanotubes in polypropylene, *Polymer* 116 (2017) 160–172.
- [15] X. Shi, G. Zhang, T. Phuong, A. Lazzeri, Synergistic effects of nucleating agents and plasticizers on the crystallization behavior of poly(lactic acid), *Molecules* 20 (1) (2015) 1579–1593.
- [16] J. Njuguna, K. Pielichowski, S. Desai, Nanofiller-reinforced polymer nanocomposites, *Polym. Adv. Technol.* 19 (8) (2008) 947–959.
- [17] M. Šupová, G.S. Martynková, K. Barabaszová, Effect of nanofillers dispersion in polymer matrices: a review, *Sci. Adv. Mater.* 3 (2011) 1–25.

- [18] M. Nurazzi Norizan, M. Harussani Moklis, S.Z.N. Demon, N. Abdul Halim, A. Samsuri, I. Syakir Mohamad, V. Feizal Knight, N. Abdullah, Carbon nanotubes: functionalisation and their application in chemical sensors, *RSC Adv.* 10 (2020) 43704–43732.
- [19] R.J. Crawford, J.L. Throne, Rotational molding polymers, in: R.J. Crawford, J.L. Throne (Eds.), *Rotational Molding Technol.*, William Andrew Publishing, Norwich, NY, 2002, pp. 19–68.
- [20] S. Park, J.O. Baker, M.E. Himmel, P.A. Parilla, D.K. Johnson, Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance, *Biotechnol. Biofuels* 3 (2010) 10.
- [21] D.S. Chaudhary, R. Prasad, R.K. Gupta, S.N. Bhattacharya, Clay intercalation and influence on crystallinity of EVA-based clay nanocomposites, *Thermochim. Acta.* 433 (1–2) (2005) 187–195, <https://doi.org/10.1016/j.tca.2005.02.031>.
- [22] G. George, M. Selvakumar, A. Mahendran, S. Anandhan, Structure–property relationship of halloysite nanotubes/ethylene–vinyl acetate–carbon monoxide terpolymer nanocomposites, *J. Thermoplast. Compos. Mater.* 30 (1) (2017) 121–140.
- [23] A. Zubkiewicz, A. Szymczyk, S. Paszkiewicz, R. Jędrzejewski, E. Piesowicz, J. Siemiński, Ethylene vinyl acetate copolymer/halloysite nanotubes nanocomposites with enhanced mechanical and thermal properties, *J. Appl. Polym. Sci.* 137 (38) (2020) 49135, <https://doi.org/10.1002/app.v137.3810.1002/app.49135>.
- [24] G. George, A. Mahendran, S. Anandhan, Use of nano-ATH as a multi-functional additive for poly(ethylene-co-vinyl acetate-co-carbon monoxide), *Polym. Bull.* 71 (8) (2014) 2081–2102.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/352993955>

In-Line and Cross-Flow Response Interactions during Vortex Induced Vibration of Marine Risers

Article in MER - Marine Engineers Review · July 2021

CITATIONS

0

READS

96

2 authors:



Vidya Chandran

SCMS School of Engineering and Technology, Karukutty

24 PUBLICATIONS 31 CITATIONS

SEE PROFILE



Sheeja Janardhanan

Indian Maritime University

66 PUBLICATIONS 77 CITATIONS

SEE PROFILE

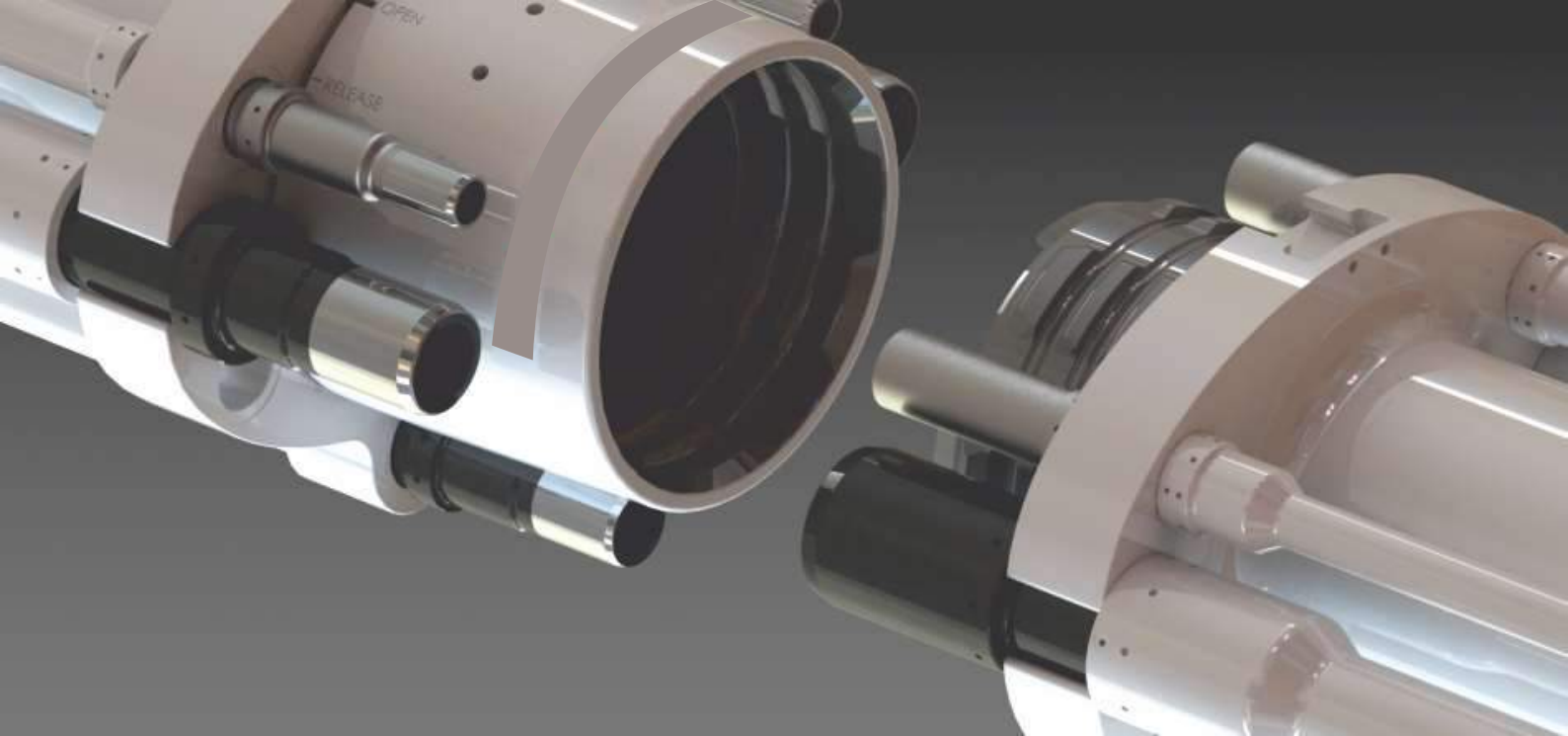
Some of the authors of this publication are also working on these related projects:



Seaplanes [View project](#)



Bio-Inspired Propulsion Systems [View project](#)



In-Line and Cross-Flow Response Interactions during Vortex Induced Vibration of Marine Risers

Vidya Chandran¹, Sheeja Janardhanan²

¹Department of Mechanical Engineering, SCMS School of Engineering and Technology, Karukutty, Kerala, India

²School of Naval Architecture and Ocean Engineering, Indian Maritime University, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

The paper presents a simplified method for understanding the interaction between in-line and cross-flow responses using computational fluid dynamics simulations. Interaction between the responses in the in-line and cross-flow directions in vortex induced vibrations of cylindrical risers in the marine environment is still not fully understood. The trends of variation of hydrodynamic and structural parameters as well as pattern of shedding have been determined numerically to understand the effect of the in-line degree of freedom as well on the riser response and hydrodynamic force coefficients and the results show that a single degree of freedom riser is more susceptible to lock in vibration.

KEYWORDS: Vortex Induced Vibration, In-line, Cross-Flow, Force Coefficients, Response

1. INTRODUCTION

Drilling riser is a pipe laid vertically from the oil well at the ocean bed to the offshore drilling platform. It conveys the drilling fluid and mud to and from the drill site. Marine drilling risers are used especially with floating rigs which are less stable and in particular cases where disconnection of the platform from the seafloor may be required quite often. **Figure 1** shows various layouts of marine risers depending on the constructional specification of platforms. **Figure 2** shows different

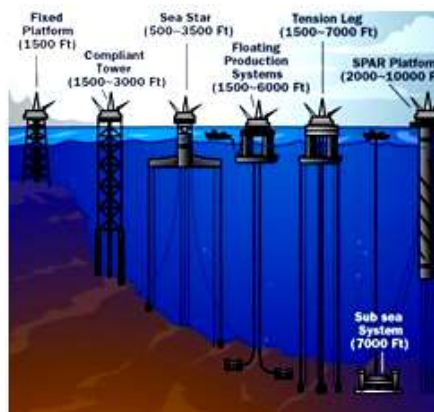


Figure 1 Constructional variation of offshore platforms with drilling depth [1]

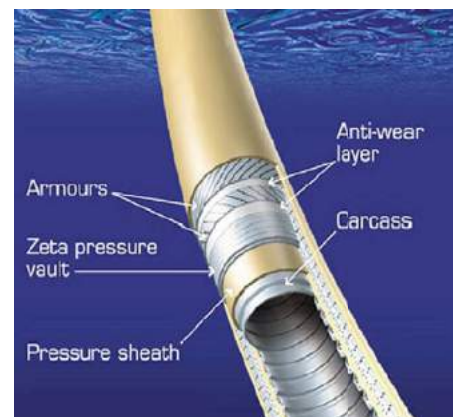


Figure 2 Cross section of a typical flexible riser [1]

Marine drilling risers are used especially with floating rigs which are less stable and in particular cases where disconnection of the platform from the sea floor may be required quite often

layers used in the construction of a flexible riser.

The marine risers, which are classified as **bluff bodies**, and when encountering fluid flow, alternate vortices are shed in the wake of the structure due to boundary layer separation. This alternate vortex shedding happens at a discrete frequency depending upon the flow Reynolds Number (Re). When the vortex shedding frequency matches with the natural frequency of the riser structure, it resonates with high amplitude of oscillation. These large amplitude vibrations, that occurs during "lock-in" of risers are catastrophic and needs to be arrested for the safety of crew working on the floating platforms and also for extending operational life of the risers. Vortex induced vibration (VIV) of marine risers poses all the challenges in the deployment and operation of marine risers.

There have been lot of research in the recent past to understand their behaviour under various sub-sea flow conditions. But most of the studies have concentrated on understanding the wake characteristics and estimating hydrodynamic loading and response of either stationary cylinder or cylinder with a single degree of freedom (1DOF)[2].

Few results have been reported for study of hydrodynamic response of cylinder with two degrees of freedom (2DOF) in both in-line (IL) and cross-flow (CF) directions. IL vibrations have significant impact on the shedding pattern and also on the amplitude of CF vibrations [3].

The first of its kind discussions were reported in the case of flow around cylinder with 2DOF [4]. They established the effect of reduced velocity (U_r) on the effect of forced and free 2DOF response [4]. The effect of IL response on CF response depends on the ratio of natural frequencies in both the directions

$$(\eta_b = \frac{f_{ILCF}}{f_{NL}})$$

During lock in, if the natural frequency in the IL direction is twice that in the CF direction, resonance occurs in both directions leading to premature failure of the riser [5]. Also it has been observed that IL response amplitude is a function of U_r and stability parameter, whereas the CF response amplitude is a function of U_r and flow velocity [6]. Wake characteristics, hydrodynamic force coefficients and response vary significantly when both IL and CF vibrations occur simultaneously. Hence there is a need for prediction of response that hold good for the combined IL and CF vibration.

2. PROBLEM DESCRIPTION

In the present paper a riser model with outer diameter 0.076 m has been numerically analysed using two dimensional (2D) computational fluid dynamics (CFD). Specifications of the riser and the flow condition in listed in **Table 1**. The incoming flow velocity is fixed as 0.5 m/s to maintain the flow regime uniform at $Re = 3.8 \times 10^4$ which corresponds to the ocean condition encountered by a real marine riser used for petroleum extraction in offshore industries [7]. In this paper an effort has been made to study the effect of IL vibration on the amplitude of CF vibration and also on the wake characteristics.

2.1. Mathematical Model

The riser has been modelled as a 2D cylinder with 2DOF in the CF and IL directions. The equations of motion for the riser can be represented as Eq. (1) and (2)

$$m\ddot{Y} + c\dot{Y} + kY = F_L(t) \quad (1)$$

$$m\ddot{X} + c\dot{X} + kX = F_D(t) \quad (2)$$

Where Y is the displacement in CF direction and X is the displacement in

Properties	Values	Units
Diameter (D)	0.076	m
Aspect ratio (L/D)	13.12	-
Flow velocity (V)	0.5	m/s
Reynolds Number of flow (Re)	3.8×10^4	-
Mass ratio (m')	0.66	-

Table 1 Riser model specifications and flow characteristics

the IL direction. The excitation forces are lift force, $F_L(t)$ and drag force, $F_D(t)$. The excitation forces are periodic in nature due the alternate shedding of vortices, which causes the riser to oscillate in CF as well as IL directions. The riser is observed to oscillate with frequency equal to frequency of vortex shedding (f_v) in the CF direction and at double the frequency in the IL direction during lock in. Lock in can be defined as the resonance condition during which the vortex shedding frequency lock on to the natural frequency of the riser in the cross flow direction. A simple representation of the mathematical model of riser with 2DOF is represented in **Figure 3**.

The riser is modelled with zero structural damping in the CF and IL directions. k_x and k_y are stiffness coefficients in the IL and CF directions respectively. In the present study $k_x = k_y$. For such a specific case the natural frequencies in both directions will be same and hence $\eta_b = 1$

2.2. Fluid Domain Extends

Figure 4 (a) shows the computational domain for the CFD simulation of VIV of an elastically mounted cylinder with 2DOF. The origin of the Cartesian coordinate system is located at the centre of the cylinder. The length of the

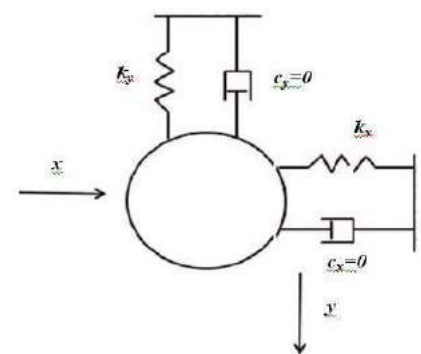


Figure 3 Representation of mathematical model of riser with 2DOF

domain is $40D$ with the cylinder located at $10D$ away from the inlet boundary. The cross flow width of the domain is $20D$ with the centre of the cylinder at the middle.

Detailed views of the mesh around the cylinder along with the computational domain after meshing have been shown in Figure 4 (b) and (c) respectively. There are 307 nodes around the circumference of the cylinder and the minimum element size near the rigid wall boundary has been computed from boundary layer theory to be $0.0001D$ [8].

The non-dimensional element size represented as y^+ , next to the cylinder surface is found to be less than unity. For cylinder wall a no slip boundary condition has been applied assuming the surface to be smooth. Inlet boundary has been treated as velocity-inlet with inflow velocity, $V = 0.5$ m/s. Outlet boundary has been treated as pressure outlet, the gradients of fluid velocity are set to zero and the pressure with zero reference pressure. On the two transverse boundaries, symmetry boundary condition has been applied. Grid independency study has been carried out for the present grid in the previous work done by the authors [9].

2.3. Flow Model

Numerically this problem has been treated as a case of two-way fluid structure interactions (2way FSI). Modeling and meshing has been performed in ANSYS ICEM CFD and solving using ANSYS FLUENT. Flow around the cylinder is modeled using the transient, incompressible Reynolds Averaged Navier-Stokes equation (RANSE) based solver with $k-\omega$ SST as the turbulence model. RANSE solver does the virtual averaging of velocities over an interval of time and hence for a specific interval, the velocity vector appears to be constant in a RANSE solver. In the present work an optimised fine grid is used to compensate for this drawback of the solver enabling it to capture the physics of Von-Karman Street eddies.

The governing equations are discretised using finite difference method. Non iterative time advancement (NITA) scheme with fractional time stepping method (FSM) has been chosen for pressure-velocity coupling of the grid. A least-squares- cell (LSC) based scheme has been used for gradient in spatial discretisation and a second order upwind scheme as convective scheme.

2.4. Structural Model

An elastically mounted cylinder can be mathematically represented by Eq. (1) and (2). These equations of motion are solved using a six degrees of freedom solver (6DOF), an integral part of the main solver by defining the cylinder as an object with 2DOF in transverse direction. A user defined function (UDF) compiled in C programming language has been hooked to the cylinder dynamic boundary conditions. The governing equations for the motion of the centre of gravity of the cylinder in the CF and IL directions are solved in the inertial coordinate system. Velocity in the CF and IL directions are obtained by performing integration on Eq. (3) and (4).

$$\ddot{y} = \frac{1}{m} \sum F_L \quad (3)$$

$$\ddot{x} = \frac{1}{m} \sum F_D \quad (4)$$

Where \ddot{x} and \ddot{y} , are accelerations in the IL and CF direction respectively, m is the mass of the cylinder and F , resultant fluid force acting on the cylinder in the respective direction. Position of the centre of gravity of the cylinder (CG) is updated after solving the equations of motion of a spring mass system represented by Eq. (1) and (2). Mass of the cylinder is given in the UDF as in Eq. (5) and (6).

$$m = m_b + m_a \quad (5)$$

$$m_a = (1 + C_A)m_b \quad (6)$$

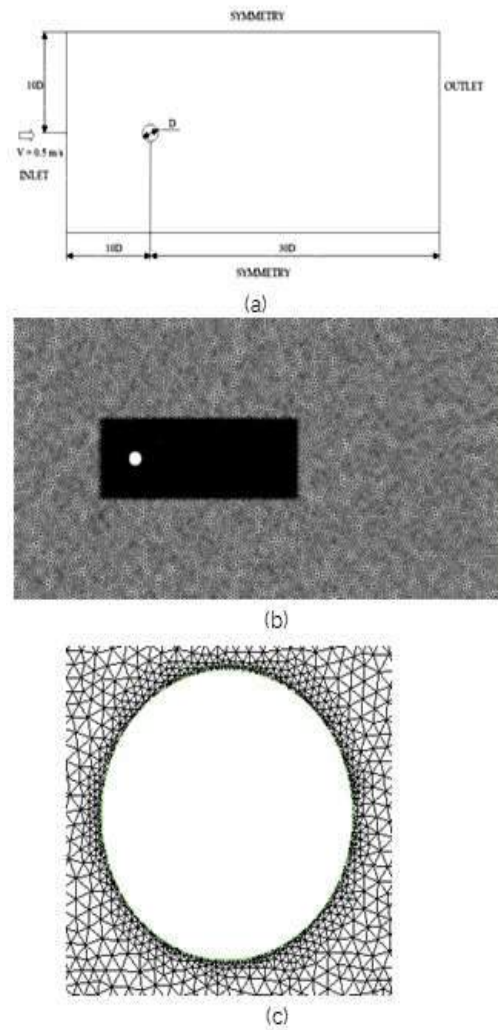


Figure 4 (a) Computational domain
(b) computational mesh
(c) mesh around the cylinder

Where m_a is the added mass and m_b is the mass of the cylinder. Added mass coefficient C_A for the aspect ratio of the present model is found to be equal to 1.0 [10].

Analysis has been performed assigning the cylinder 2DOF with $k_x = k_y$ so that the natural frequencies of the cylinder in both directions remain equal. The results are compared with the case when the cylinder has only 1DOF in the CF direction. Amplitudes of CF response are compared with existing results [9] and also the shedding patterns in both cases have been analysed.

3. RESULTS AND DISCUSSIONS

From the numerical analysis of cylinder with 2DOF it has been observed that the hydrodynamic force coefficient in the CF direction, C_L shows an increase of 17.4% than that for 1DOF case. This result is comparable with the findings of previous research in the field which shows an increase in the lift coefficient value by permitting an extra degree of freedom [11].

RMS value of C_D is constant for both cases with a very small decrease of 4% with 2DOF case. C_L oscillates about zero with almost equal frequencies for both the cases. But the frequency of oscillation of C_D is lesser by 7.2% for 2DOF case. The values of important hydrodynamic and structural parameters of both cases are shown in **Table 2**.

The non-dimensional amplitude in the CF direction obtained with 2DOF is 11.3% more than that with 1DOF. X/D is approximately 0.2. Time histories of major parameters obtained from the 1DOF analysis are shown in **Figures 5(a) – (d)** and that for 2DOF in **Figures 6(a) – (d)**. Frequency of oscillation of the cylinder in the CF direction obtained from 1DOF case is found to be more closer to the theoretical value of vortex shedding frequency obtained from the normal value of $St = 0.2$ ($f_v=1.3$). For the 2DOF case, the frequency of oscillation deviates from the vortex shedding frequency.

For 2DOF case, the frequency of oscillation of C_L and the oscillation frequency of cylinder in the CF direction remains same. In 1DOF case, C_L oscillation frequency remains same as that in the 2DOF case, but the cylinder vibration frequency in the CF shifts towards the natural frequency of cylinder in CF direction.

In the present analysis, the natural frequency in both directions are specifically fixed to be equal to the theoretical value of vortex shedding frequency. Hence the phenomenon can be looked upon as the lock-in of vortex shedding frequency on to the natural frequency of the cylinder. It can be concluded that a cylinder with 1DOF is more prone to lock in vibration compared to that with 2DOF.

This observation can be related to the shifting of the vortex shedding pattern from two singles (2S) to two pairs (2P) mode when motion in IL direction is arrested. The shedding patterns for 1DOF and 2DOF cases are shown in **Figure 5(a)** and **Figure 6(a)** respectively. St obtained also is with the range of normal value for cylinders during lock in. Even though the values of C_D for both cases are almost same, the oscillating frequency varies significantly.

The trajectory of oscillation of cylinder in 2DOF case is represented in **Figure 7**. A clear eight figure trajectory is observed which is typical for VIV of cylinders [11]. Also it has been observed that the motion the IL direction lags behind that in CF direction by a phase angle 30° . The represented trajectory in **Figure 7** corresponds to 30° phase lag [12].

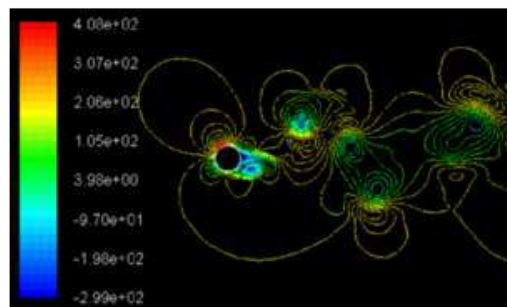
4. CONCLUSIONS

Accounting for an additional degree of freedom seems to have significant effect on the magnitude of lift coefficient but the frequency of oscillation of C_L remains constant for both the cases. C_D is independent of the degree of freedom of the cylinder but the frequency of oscillation varies significantly. Oscillation amplitude of the cylinder in the CF direction is more in 2DOF case which can be related to the increase in C_L .

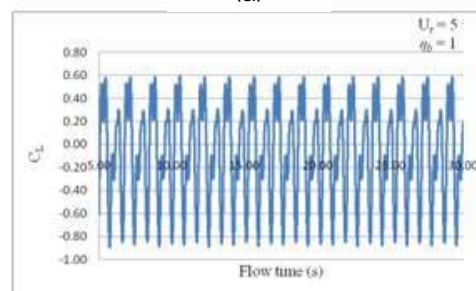
It has been clearly observed that with 1DOF, the cylinder is more susceptible to lock in vibration since the vortex shedding frequency locks on to the natural frequency of the cylinder in the CF direction. But with 2DOF, no such shifting of frequency is observed. Shedding pattern shifts from 2S during 2DOF motion to 2P when motion in IL direction is arrested. An eight figure trajectory typical for VIV is obtained from the

Parameters	1DOF	2DOF ($\eta_b = 1$)
C_L	0.57	0.69
C_D	1.49	1.43
$f_{osc C_L}(f_v)$	1.16	1.14
$f_{osc C_D}$	2.5	2.32
f_{CF}	1.26	1.15
f_{IL}	-	3.28
St	0.18	0.17
Y/D	1.06	1.2
X/D	-	0.17

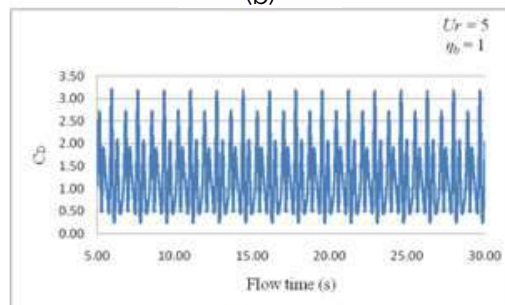
Table 2 Hydrodynamic and structural parameter off cylinder with 1DOF and 2DOF



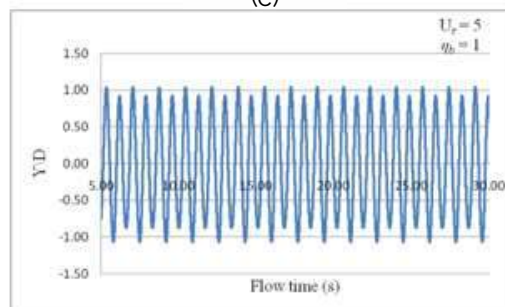
(a)



(b)



(c)



(d)

Figure 5 Pressure contours and Time histories of various hydrodynamic and structural parameters (a) Vortex shedding pattern behind cylinder with 1DOF showing 2P mode (b) C_L of cylinder with 1DOF (c) C_D of cylinder with 1DOF (d) Motion history of cylinder with 1DOF

It can be concluded that a cylinder with 1DOF is more prone to lock in vibration compared to that with 2DOF

2D simulation. Hence the efficacy of 2D CFD as a tool to predict response of cylinder with 2DOF under VIV is accomplished. The observations made above are definitely strong inputs in the design deployment and operation of marine risers.

5. REFERENCES

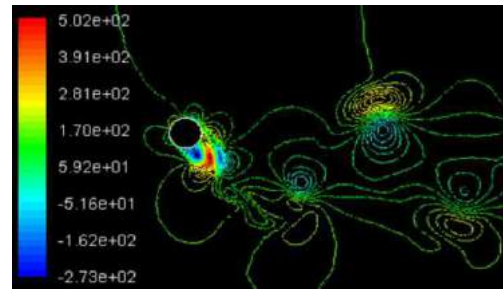
1. Bai, Y., and Qiang B., *Subsea engineering handbook*. Gulf Professional Publishing, Vol. 1, pp. 28, 2018.
2. A. Khalak, C. H. K. Williamson, 'Investigation of the relative effects of mass and damping in vortex induced vibration of a circular cylinder', *Journal of Wind Engineering. Ind. Aerodyn.* vol. 69-71, pp. 341 – 350, 1997.
3. Jauvtis, N. and Williamson, C. H. K., 'The effect of two degrees of freedom on vortex-induced vibration at low mass and damping', *Journal of Fluid Mechanics* 509, 23 – 62, 2004.
4. Moe, G. and Wu, Z. J., 'The lift force on a cylinder vibrating in a current', *Journal of Offshore Mechanics and Arctic Engineering* 112, 297- 303, 1990.
5. T. Sarpkaya., 'Hydrodynamic damping, flow-induced oscillations, and biharmonic response', *ASME Journal of Offshore Mechanics and Arctic Engineering*, 117:232-238, 1995.
6. Yin, D., Experimental and Numerical Analysis of Combined In-line and Cross-flow Vortex Induced Vibrations. *Ph D Thesis, Norwegian University of Science and Technology* pp 7, 2013.
7. Narendran, K., Murali, K., Sundar, V., 'Vortex-induced vibrations of elastically mounted circular cylinder at Re of the O (10^5)', *Journal of Fluids and Structures*, 54, 503 – 521, 2015.
8. Schlichting, H., 'Boundary layer theory', *McGraw-Hill Book Company, New York*, 1979.
9. Chandran, V., Sekar, M., Janardhanan, S., Menon, V., 'A numerical study on the influence of mass and stiffness ratios on the vortex induced motion of an elastically mounted cylinder for harnessing power', *Energies*, 11, 2580, 2018.
10. Naudascher, E.; Rockwell, D., 'Flow induced vibration – An engineering guide', *Dover Publications Inc, Mineola, New York, USA*, 2005.
11. Williamson, C.H.K., Govardhan, R., 'Vortex induced vibrations' *Annual review of fluid mechanics*, vol. 36 pp 413 – 455, 2004.
12. W, Jie., L, Halvor., M, L, Larsen., L, Stergios., B, Rolf. 'Vortex-induced vibration of a flexible cylinder: Interaction of the in-line and cross-flow responses', *Journal of Fluids and Structures* 63 238–258, 2016.

ABOUT THE AUTHORS

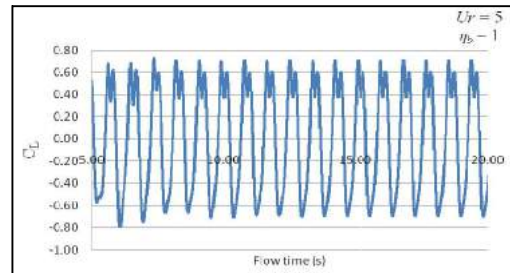
Dr. Vidya Chandran is Associate Professor in the Department of Mechanical Engineering, SCMS School of Engineering and Technology, Cochin. She has a PhD in Vortex Induced Vibrations from Karunya University, India. Her research interests include Vortex Induced Vibrations, Marine Clean Energy and Under Water Robotics.

Dr. Sheeja Janardhanan is Associate Professor in the School of Naval Architecture and Ocean Engineering, Indian Maritime University, Vishakhapatnam, India. She has a PhD in Numerical Ship Hydrodynamics from Indian Institute of Technology Madras, India. Formerly she worked as Professor and Head, Department of Mechanical Engineering, SCMS School of Engineering and Technology, Ernakulam and also as Surveyor in the Research and Rule Development Division of Indian Register of Shipping, Mumbai, India. Her research interests include controllability of surface ships, underwater robotics, vibrations of risers and computational fluid dynamics.

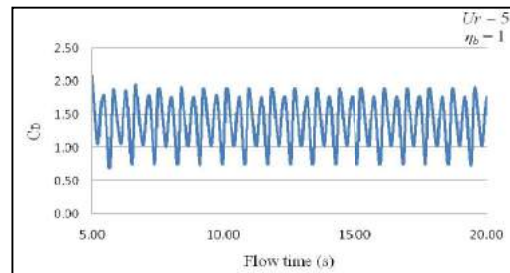
Email: sheeja@imu.ac.in



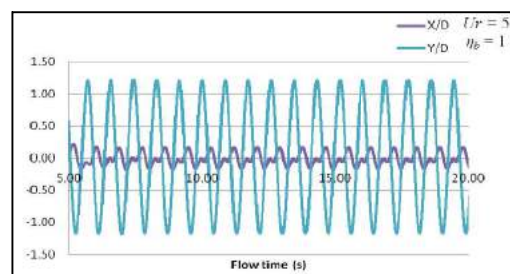
(a)



(b)



(c)



(d)

Figure 6 Pressure contours and time histories of various hydrodynamic and structural parameters (a) Vortex shedding pattern behind cylinder with 2DOF showing 2S mode (b) C_L of cylinder with 2DOF (c) C_D of cylinder with 2DOF (d) Motion history of cylinder with 2DOF

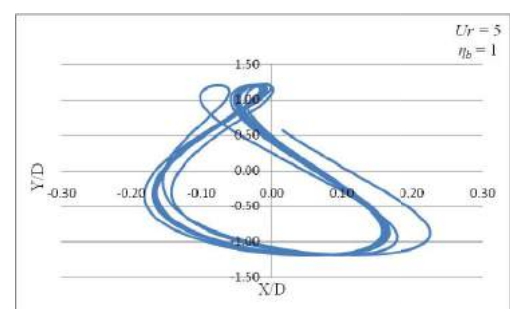


Figure 7 Trajectory of the cylinder with 2DOF motions under VIV

PAPER • OPEN ACCESS

Impact of Ground Nut Shell Ash on Cobalt-Chromium metal matrix composites synthesized using Powder metallurgy process.

To cite this article: G R Raghav *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1166** 012006

View the [article online](#) for updates and enhancements.

You may also like

- [High-rate multi-GNSS attitude determination: experiments, comparisons with inertial measurement units and applications of GNSS rotational seismology to the 2011 Tohoku Mw9.0 earthquake](#)
Peiliang Xu, Yuanming Shu, Xiaoji Niu et al.

- [Using Allan variance to evaluate the relative accuracy on different time scales of GNSS/INS systems](#)
Quan Zhang, Xiaoji Niu, Qijin Chen et al.

- [Investigation on mechanical, wear and corrosion properties of Fe-Co-Cr-W-GNSA hybrid composites synthesized using powder metallurgy process](#)
G R Raghav, D MuthuKrishnan, R Sundar et al.



Connect with decision-makers at ECS

Accelerate sales with ECS exhibits, sponsorships, and advertising!

▶ Learn more and engage at the 244th ECS Meeting!

Impact of Ground Nut Shell Ash on Cobalt-Chromium metal matrix composites synthesized using Powder metallurgy process.

G R Raghav¹, Suraj R^{2*}, Sheeja Janardhanan³, Vidya Chandran⁴,
K J Nagarajan⁵ and Nikhil Asok⁶

^{1,2,3,4,6}Department of Mechanical Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala, India.

⁵Department of Mechanical Engineering, KLN College of Engineering, Pottapalayam, Sivagangai Dt. Tamil Nadu, India 630612

*Corresponding Author email id: surajr@scmsgroup.org

Abstract. Co-based composites are extensively utilized in the field of prosthesis and dental implants. Hybrid composites made using Powder metallurgy process, Co-10Cr-GNSA were studied. The surface morphology of the hybrid composites were studied using Scanning Electron Microscope. The elemental analysis was carried out using X-Ray Diffraction technique. The hybrid composites were analyzed for its various mechanical properties like microhardness, compressive strength, and density. Value of micro hardness of the composite materials showed slight improvement with addition of GNSA reinforcement. The value of density of the hybrid composites was found to be decreasing linearly with the addition of GNSA. Compressive strength of the materials showed a reasonable increment. Wear analysis to study the tribological characterization of the hybrid composites were done with the help of a pin on disc wear testing machine. The wear and COF studies show that with a rise in GNSA content, wear resistance increases because of the presence of oxides of GNSA particles. From the worn out surfaces of the hybrid composite it is concluded that the deformation of the composites takes places initially due to abrasive wear followed by plastic deformation. An electrochemical workstation was used to understand the corrosion characteristics of the hybrid composites in the presence of 3% NaClelectrolyticsolution. Co-5Cr-5GNSA hybrid composites exhibit better electrochemical corrosion resistance compared to other specimens.

Keywords: Powder metallurgy, Wear, Corrosion, GNSA

1. Introduction

Now a days more and more people suffer from osteoarthritis disorder, which makes them experience severe pain and discomfort. Recent survey suggests that there are nearly 50 million cases worldwide who are suffering from osteoarthritis disorder and in need of joint replacement surgery [1]. Co-Cr-Mo alloy is the extensively used artificial prosthetic material considering its higher value of wear, hardness and corrosion resistance. Even though Co-Cr-Mo alloys are excellent prosthetic material, still there are certain disadvantages such as wear of implants in the hip joints and problems related to bio compatibility since Mo is not a bio degradable material[2–5]. Therefore it is the need of the hour to



produce a composite with much better wear and corrosion resistance which is also bio degradable and compatible to human body.

The ground Nut Shell Ash (GNSA) which is primarily a biological waste and is available in abundance all over the world. Moreover the GNSA particles have presence of $MgSiO_3$ and $AlSiO_3$ in high concentration. Hence it can be used to replace the hazardous Mo reinforcements[6].

There are many conventional methods to produce wear resistance artificial prosthetic implants such as plasma spraying, physical vapor deposition, electro deposition and chemical vapour deposition. Since these manufacturing processes includes more complex steps and requires costly equipments, the cost of the implants is high. The powder metallurgy technique has its own advantages which include uniform dispersion, low processing cost and ability to manufacture high melting point materials. Hence the powder metallurgy process possesses great potential for producing Co-Cr based hybrid composite materials with highly desirable mechanical properties along with wear and corrosion resistance[7–13].

This work aims to develop a Co-Cr-GNSA hybrid composite material with better wear, corrosion resistance and mechanical properties. In this study, four different compositions based on weight percent is formulated as follows Co-10Cr, Co-10Cr-2.5GNSA, Co-10Cr- 3.5 GNSA and Co-10Cr-5GNSA. The composite powders are mechanically milled and compacted and sintered in order to develop specimens of 8mm cylindrical pellets. The hybrid composites are then studied in order to explore their morphological properties using SEM. The mechanical behavior along with tribological and corrosion resistance behavior were studied and their mechanisms were reported.

2. Materials and Method

The materials CoCr (99.5% purity) which is used in this study were purchased from Mepco Ltd Tamil Nadu, India. The ground nut shell ash (GNSA) powder used in this work is prepared using heat treatment method which is discussed in our pervious paper [6]. Mechanical ball milling process was used for alloying the Co-Cr- GNSA hybrid composites. The process was carried out for two hours and was then compacted into 8 mm diameter pellet which is cylindrical in shape. The value of compaction pressure was set to 750 MPa consistently. After this, the soft green compacts were hardened by forcing them to sintering process at 1000oC for 2h. The morphology of the hybrid composites were studied using a Field Emission Scanning Electron Microscope (FE-SEM). ASTM: B962-13 standards were used to calculate the density of the Co-Cr- GNSA hybrid composites. The ASTM E384 standards were used to study the micro hardness of the hybrid composite pellets at a uniform load and dwell time of 1 kgf and 10 seconds respectively. Compressive strength of the hybrid composites were studied at a scan rate of 5 mm/min, with the help of a Universal Testing Machine (UTM). ASTM G99-05 standards were used to study the wear and friction behavior of the composites. EN 32 steel of hardness 65 HRC was used for the analysis. The specimens were cleaned using acetone solution before and after the wear test. The wear analysis of the composites was done at various sliding conditions such as the load, sliding distance and sliding speed. The electrochemical corrosion tests were simulated on a three electrode workstation using 3% NaCl solution as electrolyte[14–16].

3. Results and Discussion

3.1 Field Emission Scanning Electron Microscope Analysis

FE-SEM images of Co-10Cr- 3.5 GNSA & Co-10Cr-5GNSA hybrid Composites respectively are shown in Figure 1. There is a homogenous mixture of GNSA Particles with Co and Cr particles. The wettability of the GNSA particles was the major factor in achieving uniform amalgamation. It can be noted that due the milling operation the size of Cr particles have reduced to around 500 nm in size and are bonded strongly with Co matrix.

3.2 Microhardness

The microhardness test was done using a Vickers Micro Hardness Testing Machine with the test being conducted at five different points. Figure 2 shows the variation in the average value of microhardness

of the composites at different configurations based on its composition, i.e. Co-10Cr, Co-10Cr-2.5GNSA, Co-10Cr- 3.5 GNSA and Co-10Cr-5GNSA. The microhardness of the composites varied from 320 HV to 340 HV. The hardness of Co-10Cr was found to be 320 HV and the introduction of GNSA resulted in an increase in the microhardness. The maximum microhardness was found to be in Co-10Cr-5GNSA composite with a value of 340 HV. The uniform amalgamation of GNSA particles was the major reason for this improvement in microhardness.

3.3 Compressive Strength and Density

With the addition of the GNSA reinforcement the density of the Co-10Cr –GNSA hybrid composites were found to be decreasing. The value of density for Co-10Cr composites was recognized as 8.1 g/cm³ whereas the density of the Co-10Cr-5GNSA hybrid composites were around 7.65 g/cm³ as shown in Figure.3. This reduction in density was attributed by the relatively soft nature of the GNSA particles. With the addition of GNSA particles, the compressive strength of the hybrid composite materials showed slight increase in its value. Figure.3 helps us understand the compressive strength of different combinations of Co-10Cr-GNSA hybrid composites. The compressive strength of Co-10Cr composite was established to be in the region of 380 MPa. The compressive strength has slightly increased to 401 MPa for the Co-10Cr- 5 GNSA hybrid composites which is due the presence of AlSiO₃ particles in the GNSA ash content.

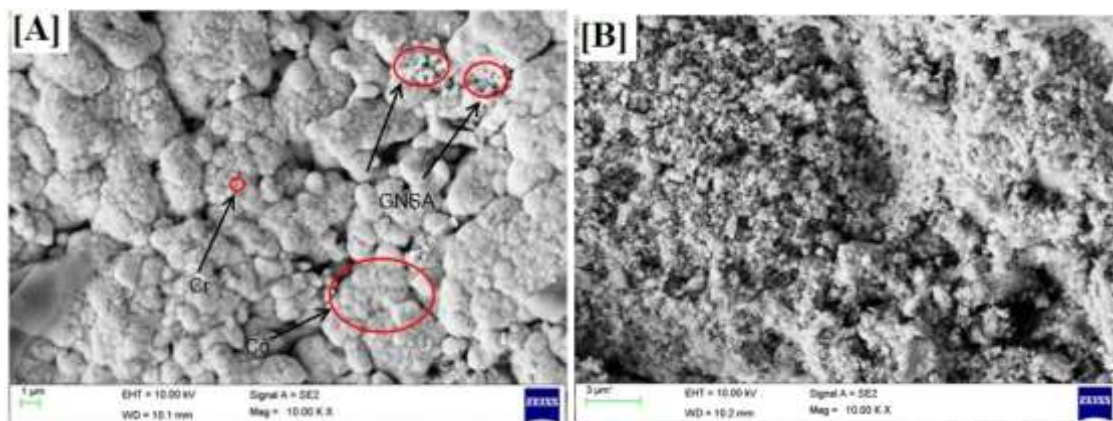


Figure 1.FESEM images of Co-10Cr- 3.5 GNSA & Co-10Cr-5GNSA hybrid Composite.

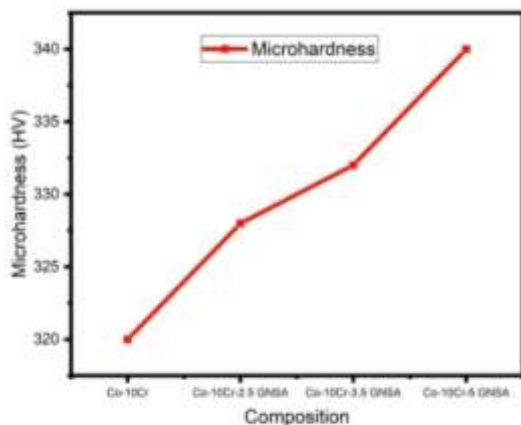


Figure 2.Graphical Representation of Co-10Cr-GNSA hybrid composites.

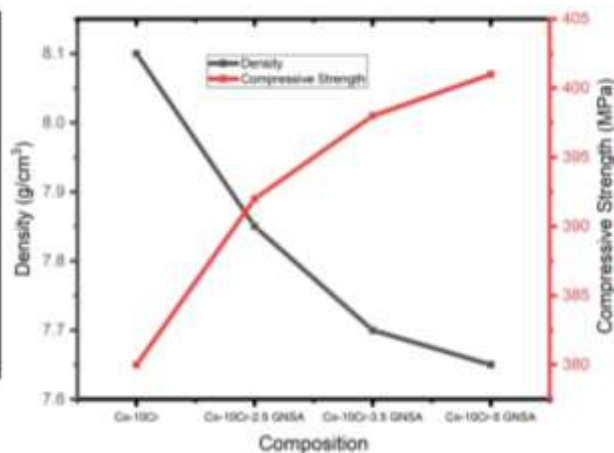


Figure 3.Comparison of Density and Compressive Strength of the Co-10Cr-GNSA hybrid composites.

3.4 Wear and COF Analysis

The loss of material due to wear of the Co-10Cr-GNSA hybrid composites is shown in Figure 4. The variation of wear loss of Co-10Cr-GNSA hybrid composites is depicted as graphical plots. The Figure 4 (A) indicates the wear analysis data of the Co-10Cr-GNSA hybrid composites at different loads (10N, 15N and 20N). The sliding speed (1.5 m/s) and sliding distance (1000 m) were kept constant. The Co-10Cr-5GNSA hybrid composites have witnessed very minimal wear loss at all loading conditions. The wear loss of Co-10Cr-GNSA hybrid composites at various sliding distance and speed is shown in Figure 4 (B&C) respectively. The wear loss has experienced similar trend. With the increase in GNSA concentration in the matrix there is definite resistance to wear and thereby the wear loss is very minimal for the Co-10Cr-5GNSA hybrid composites. The variation in coefficient of friction at different loads, sliding distance and sliding speed for Co-10Cr-GNSA hybrid composites is depicted in Figure 5(A,B&C). It was observed that with an increase in load, the COF of the hybrid composites increased. Whereas, it reduced with an increase in sliding speed. Overall the Co-10Cr-5GNSA hybrid composites displayed better COF value. This improvement in Wear and friction characteristics is may be attributed to the presence of AlSiO₃ compounds in the composite material and also due to the tribo oxide surface layer formation on the surface of the composite specimen. The worn out surface analysis of the Co-10Cr-GNSA hybrid composites after wear analysis is represented in Figure 6. From the worn out surface analysis it can be concluded that there is plastic deformation experienced in hybrid composites which is preceded by abrasive wear.

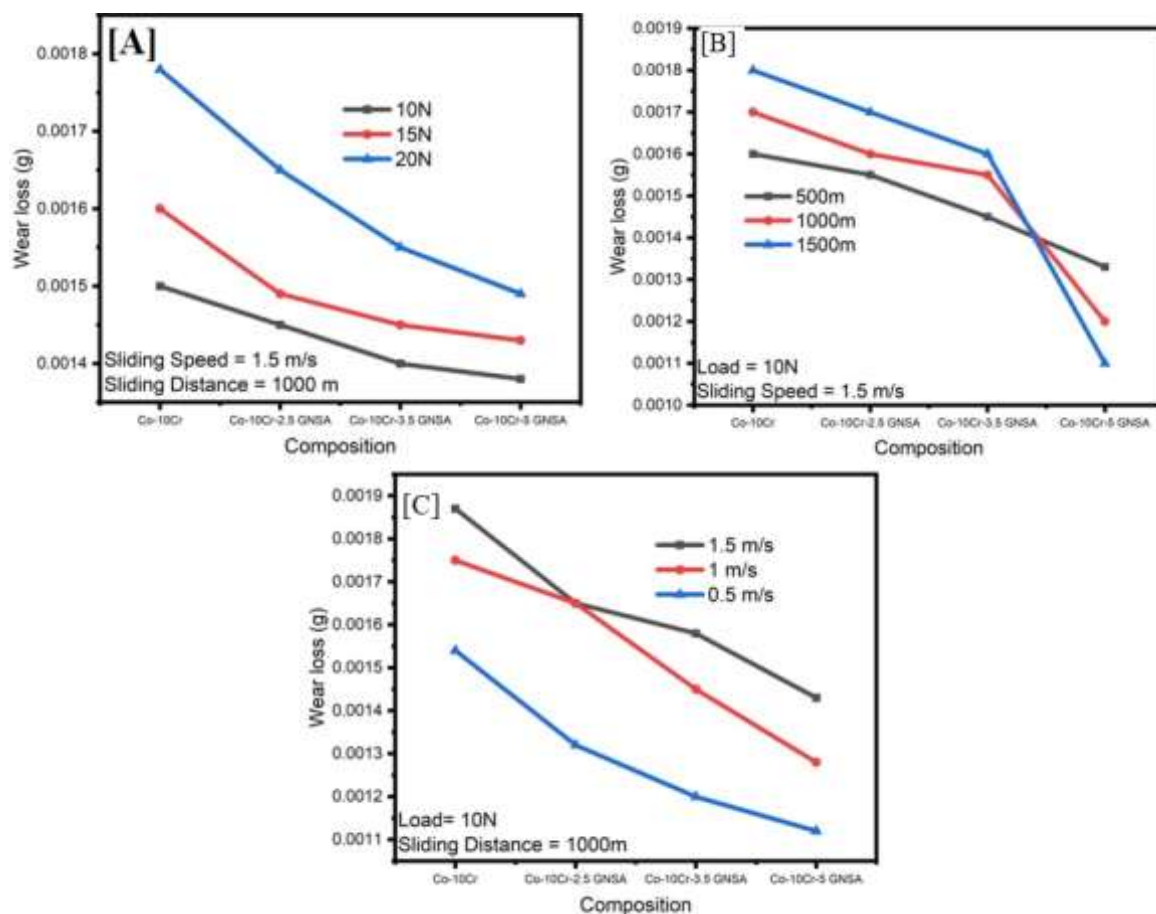


Figure 4. Wear Loss plot of Co-10Cr-GNSA hybrid composites.

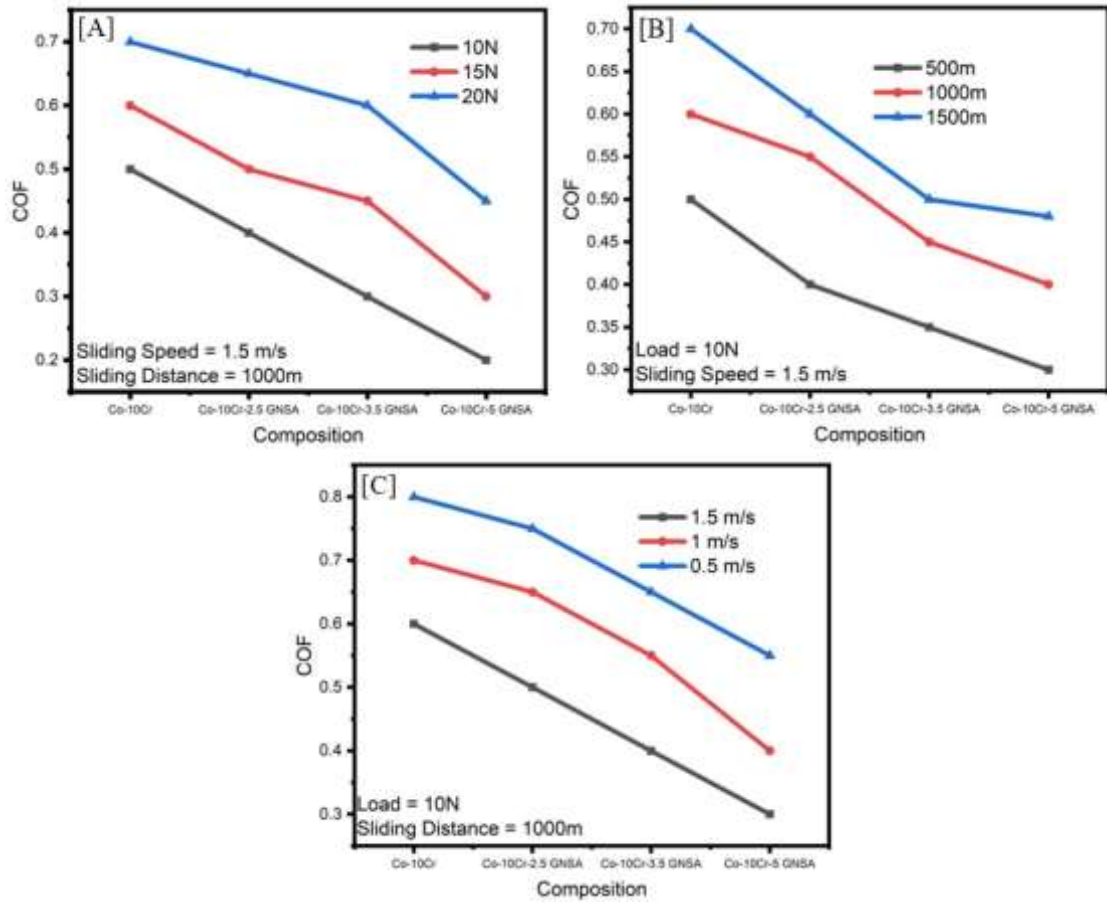
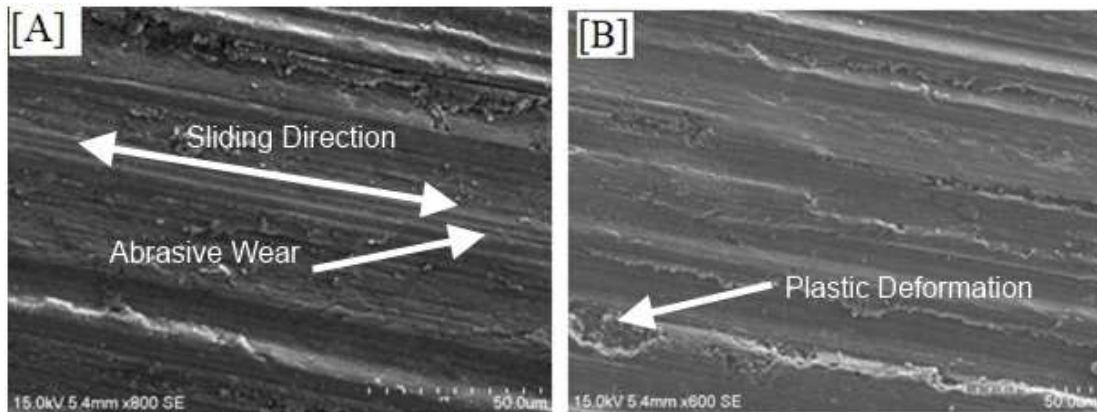


Figure 5. COF plot of Co-10Cr-GNSA hybrid composites.



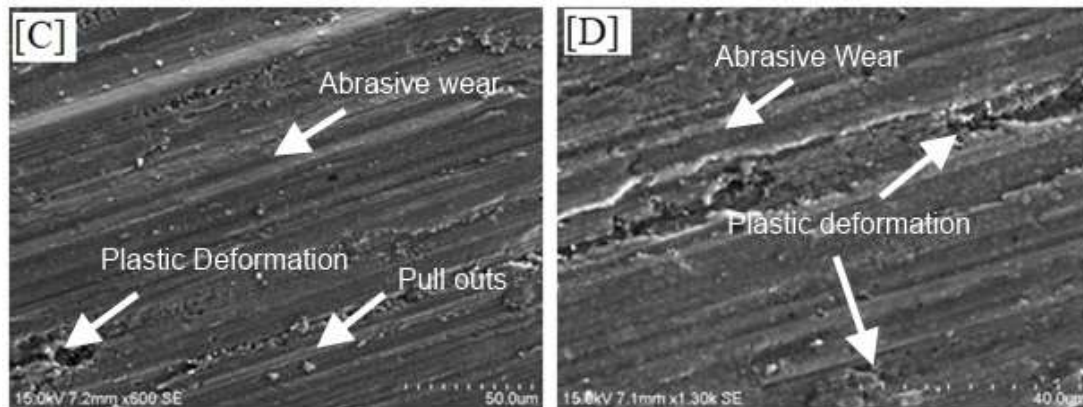


Figure 6.Worn out Surface analysis of Co-10Cr-GNSA hybrid composites.

3.5 Electrochemical Corrosion Analysis.

The corrosion analyses of the Co-10Cr-GNSA hybrid composites were done using an electrochemical work station with three electrodes. The electrolyte which was used in this study is 3% NaCl solution. The polarization curves are obtained by using tafel extrapolation methods as shown in Figure.7. The test results exhibit that the corrosion potential value, E_{corr} and the corrosion current value, I_{corr} of Co-10Cr-5GNSA hybrid composites was found to be better compared to other combinations of hybrid composites. The E_{corr} value of Co-10Cr-5GNSA hybrid composites was found to be -0.419 V and I_{corr} value was around -0.12 mA/cm². The corrosion performance of Co-10Cr-3.5 GNSA was also similar to that of Co-10Cr-5GNSA hybrid composites. The Co-10Cr composite shows lesser corrosion resistance than the hybrid composites as shown in Table.1.

Table 1.Tafel plot fallouts of Co-10Cr-GNSA hybrid composites.

S.No	Specimen	E_{corr} (V)	I_{corr} (mA/cm ²)
1	Co-10Cr	-0.442±0.051	0.5±0.020
2	Co-10Cr-2.5 GNSA	-0.437± 0.044	0.4±0.011
3	Co-10Cr-3.5GNSA	-0.420±0.021	-0.1±0.003
4	Co-10Cr-5GNSA	-0.419±0.0191	-0.1±0.002

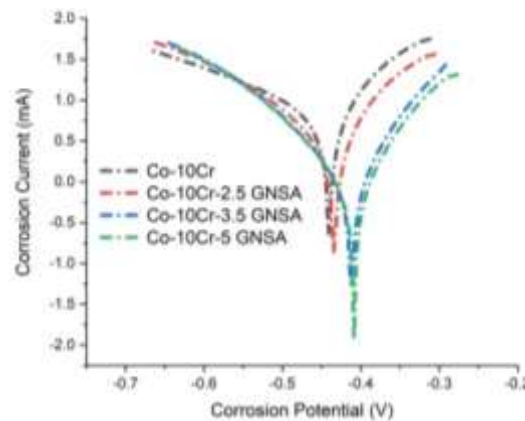


Figure 7.Potentiodynamic polarization plot of Co-10Cr-GNSA hybrid composites

4. Conclusions

The Co-10Cr-GNSA hybrid composites were studied and their mechanical, Wear and corrosion mechanisms were reported.

- The addition of GNSA reinforcement resulted in an increment in the Microhardness of the Co-10Cr-5GNSA hybrid composites (340 HV) compared to Co-10Cr composites.
- The compression strength of the Co-10Cr-5GNSA hybrid composites (401 MPa) has improved considerably than the Co-10Cr composites.
- The value of density of the Co-10Cr-5GNSA hybrid composites showed a considerable decrement due to the addition of less dense GNSA reinforcement.
- The Co-10Cr-5GNSA hybrid composites exhibited a higher resistance to wear.
- Corrosion resistance of Co-10Cr-5GNSA hybrid composites was found to be better than the Co-10Cr composites from the electrochemical corrosion analysis.

References:

- [1] Han Y, Liu F, Zhang K, Huang Q, Guo X, Wang C. *A study on tribological properties of textured Co-Cr-Mo alloy for artificial hip joints*. Int J Refract Met Hard Mater 2021;**95**:105463. <https://doi.org/https://doi.org/10.1016/j.ijrmhm.2020.105463>.
- [2] Marques FP, Scandian C, Bozzi AC, Fukumasu NK, Tschiptschin AP. *Formation of a nanocrystalline recrystallized layer during microabrasive wear of a cobalt-chromium based alloy (Co-30Cr-19Fe)*. Tribol Int 2017;**116**:105–12. <https://doi.org/10.1016/j.triboint.2017.07.006>.
- [3] Yamanaka K, Mori M, Torita Y, Chiba A. *Impact of minor alloying with C and Si on the precipitation behavior and mechanical properties of N-doped Co–Cr alloy dental castings*. Mater Sci Eng C 2018; **92**:112-120.. <https://doi.org/10.1016/j.msec.2018.06.035>.
- [4] Zhou Y, Li N, Yan J, Zeng Q. *Comparative analysis of the microstructures and mechanical properties of Co-Cr dental alloys fabricated by different methods*. J Prosthet Dent 2018:1–7. <https://doi.org/10.1016/j.prosdent.2017.11.015>.
- [5] Rodrigues WC, Broilo LR, Schaeffer L, Knörschild G, Romel F, Espinoza M. *Powder metallurgical processing of Co – 28 % Cr – 6 % Mo for dental implants : Physical , mechanical and electrochemical properties*. Powder Technol 2011;**206**:233–8. <https://doi.org/10.1016/j.powtec.2010.09.024>.
- [6] Raghav GR, Muthu Krishnan D, Sundar R, Ashokkumar R, Nagarajan KJ. *Investigation on mechanical, wear and corrosion properties of Fe-Co-Cr-W-GNSA hybrid composites synthesized using powder metallurgy process*. Eng Res Express 2020;**2**. <https://doi.org/10.1088/2631-8695/ab9517>.
- [7] Gopinath S, Prince M, Raghav GR. *Enhancing the mechanical, wear and corrosion behaviour of stir casted aluminium 6061 hybrid composites through the incorporation of boron nitride and aluminium oxide particles*. Mater Res Express 2020;**7**:016582. <https://doi.org/10.1088/2053-1591/ab6c1d>.
- [8] Prakash C, Singh S, Verma K, Sidhu SS, Singh S. *Synthesis and characterization of Mg-Zn-Mn-HA composite by spark plasma sintering process for orthopedic applications*. Vacuum 2018;**155**:578–84. <https://doi.org/10.1016/j.vacuum.2018.06.063>.
- [9] Elkhoshkhany N, Hafnway A, Khaled A. *Electrodeposition and corrosion behavior of nano-structured Ni-WC and Ni-Co-WC composite coating*. J Alloys Compd 2017;**695**:1505–14. <https://doi.org/10.1016/j.jallcom.2016.10.290>.
- [10] Stewart DA, Shipway PH, McCartney DG. *Abrasive wear behaviour of conventional and nanocomposite HVOF-sprayed WC – Co coatings*,Wear 1999;**225-229**:789–798.
- [11] Liu C, Su F, Liang J. *Nanocrystalline Co-Ni alloy coating produced with supercritical carbon dioxide assisted electrodeposition with excellent wear and corrosion resistance*. Surf Coat Technol 2016;**292**:37–43. <https://doi.org/10.1016/j.surfcoat.2016.03.027>.
- [12] Bajat JB, Vasilic R. *Corrosion Stability of Oxide Coatings Formed by Plasma Electrolytic*

- Oxidation of Aluminum : Optimization of Process Time* 2013;**69**:693–702.
- [13] Gadow R, Killinger a., Stiegler N. *Hydroxyapatite coatings for biomedical applications deposited by different thermal spray techniques*. Surf Coatings Technol 2010;**205**:1157–64. <https://doi.org/10.1016/j.surfcoat.2010.03.059>.
- [14] Raghav GR, Balaji AN, Selvakumar N, Muthukrishnan D, Sajith E. *Effect of tungsten reinforcement on mechanical, tribological and corrosion behaviour of mechanically alloyed Co-25C Cermets nanocomposites*. Mater Res Express 2019;**6**. <https://doi.org/10.1088/2053-1591/ab4f0a>.
- [15] Raghav GR, Balaji AN, Muthukrishnan D, Sruthi V, Sajith E. *An experimental investigation on wear and corrosion characteristics of Mg-Co nanocomposites*. Mater Res Express 2018;**5**:066523. <https://doi.org/10.1088/2053-1591/aac862>.
- [16] Toptan F, Alves AC, Kerti I, Ariza E, Rocha LA. *Corrosion and tribocorrosion behaviour of Al-Si-Cu-Mg alloy and its composites reinforced with B4C particles in 0.05M NaCl solution*. Wear 2013;**306**:27–35. <https://doi.org/10.1016/j.wear.2013.06.026>.

PAPER • OPEN ACCESS

Tribological and Corrosion analysis of Co-20Al-GNSA composites produced through powder metallurgy process

To cite this article: G R Raghav *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1126** 012018

View the [article online](#) for updates and enhancements.

You may also like

- [Ca-Mg Multiple Deoxidation of Ti-50Al-2Cr-2Nb Intermetallic Compound Powder for Additive Manufacturing](#)
Seongjae Cho, Taeheon Kim and Jae-Won Lim
- [Comparative study of hot corrosion behavior of thermal sprayed alumina and titanium oxide reinforced alumina coatings on boiler steel](#)
Gurdeep Singh, Santosh Kumar and Rakesh Kumar
- [Investigations into the effects of volatile biomass tar on the performance of Fe-based CLC oxygen carrier materials](#)
Matthew E Boot-Handford, Nick Florin and Paul S Fennell



Connect with decision-makers at ECS

Accelerate sales with ECS exhibits, sponsorships, and advertising!

▶ Learn more and engage at the 244th ECS Meeting!

Tribological and Corrosion analysis of Co-20Al-GNSA composites produced through powder metallurgy process

G R Raghav^{1*}, U Arunachalam², R Sujith³

¹Associate Professor, Department of Mechanical Engineering, SCMS School of Engineering and Technology, Vidyannagar, Karukutty, Ernakulam- 683576, Kerala, India

² Assistant Professor, Department of Mechanical Engineering, University College of Engineering, Nagercoil, 629004, Tamilnadu, India

³ Assistant Professor, Department of Mechanical Engineering, SCMS School of Engineering and Technology, Vidyannagar, Karukutty, Ernakulam- 683576, Kerala, India

* E-mail: raghavmechklnc@gmail.com

Abstract. In this study, Co-20Al-GNSA metal matrix composites were produced using mechanical alloying process. The Co-20Al-GNSA composites were mixed using a high-energy ball mill at a constant speed of 350 rpm for 2 hours. The composite powders were then characterized for their morphological study using Scanning Electron Microscope. The composite powders are then compressed and sintered at 500 MPa and 700°C respectively. The density and compressive strength of the composite materials shows decrement values whereas the wear resistance of the composite materials has increased considerably. The mechanism of wear was identified as abrasive wear. The electrochemical corrosion test also reveals that the Co-20Al-10GNSA composites have better corrosion resistance. The weight-loss corrosion test also shows that the composites with 10GNSA content have better corrosion resistance.

Keywords: Wear, Corrosion, Powder Metallurgy, Microhardness.

1. Introduction

The exploitation of hybrid composites for the potential replacement of conventional metals has been drastically increased in several applications such as automobile industries, commercial industrial applications, and also in aerospace industries where enhanced mechanical, wear and corrosion resistance properties are expected. Hence it was the main objective of the researchers to develop materials with lesser density with better tribological and corrosion performance [1–3].

In the process of developing a composite material, it is very important to select the matrix materials and reinforcements with the good wet ability to improve the bonding of the composite materials. Another important factor is the selection of the fabrication method and its working parameters as per the matrix and reinforcement materials. Now a day the production economy is one of the important factors due to the economic thoughtfulness of the industries [4–7]. They prefer low-cost composites, to reduce the production cost due to the raw materials. The Co is a material with very good mechanical and corrosion resistance property whereas it is very costly. So it is necessary to tailor



the mechanical properties using low-cost reinforcements such as Al and GNSA. The most generally used reinforcements are SiC, Al₂O₃, and TiO₂. The ceramic particles Al₂O₃ and TiO₂ do not have good wetting characteristics [8–10]. Thereby to improve the wettability the Ground Nut Shell Ash (GNSA) is utilized as secondary reinforcement.

The uniform dispersion of reinforcement was another area that has to be addressed while selecting a fabrication process. There are many methods such as stir casting to produce composite materials whereas achieving uniform dispersion is not possible due to cluster formation. But it is possible to produce composite materials with a uniform dispersion of reinforcements using the powder metallurgy process. Another reason for choosing powder metallurgy was their cost-effectiveness and their reliability for the production of high melting point materials [5,8,11,12].

In this study, a range of combinations of Co-20Al-GNSA hybrid composites was formed via the powder metallurgy process. The hybrid composites were then made-up into 10mm cylindrical pellets using a die setup made up of high-speed steel die. The muffle furnace was utilized to harden the compacted green pellets using a sintering operation. The microhardness, density, and compressive strength of the composite materials were studied and reported. The pin on disc apparatus and electrochemical workstation were utilized to study the wear and corrosion resistance properties respectively. Thus the main purpose of this effort is to develop hybrid composite materials with better mechanical, improved wear, and corrosion resistance properties that can be employed in automobile, industrial, and aerospace applications.

2. Materials and Methods

The chemicals used in this work are of research-grade (99.5% purity). The composite powders are synthesized using a high-energy ball mill that comprises tungsten carbide balls. The ball milling process was carried out for a duration of 2 hours at a speed of 350 rpm in the existence of toluene as a process control agent to acquire homogenous hybrid composites. The homogeneously unified composite powders are then packed down using a uniaxial hydraulic press at 500 Mpa and sintered at 700°C to produce a 10 mm cylindrical pellet. The composite materials are characterized using SEM to find out the morphology of the composite materials. The microhardness of the Co-20Al-GNSA hybrid composites was carried out at 0.5 kg load using Vickers hardness equipment and standard deviation values are considered and reported. The density of the Co-20Al-GNSA hybrid composites was calculated based on the Archimedes principle. The 10 mm diameter composite pellets are compressed using the universal testing machine (UTM) at a uniform and gradual speed rate of 3mm min⁻¹ [8,13,14]. The wear analysis was carried out at constant load, constant speed, and sliding distance of 10N, 1.5 m/s, and 1000m respectively. The electrochemical corrosion analysis was carried out using a bio-logic electrochemical workstation. The workstation consists of three electrodes, the platinum counter electrode, Al/AgCl reference electrode and composite pellets as working electrode. The scan was carried out at 5mV/s. The composite pellets are immersed in 3% NaCl solution for 1 hour so as to stabilize the open circuit potential. The weight-loss corrosion analysis was carried out with various corrosive media such as 0.1N HCl, 0.1N H₂SO₄ and 3% NaCl solution. The composite pellets are immersed in the corrosive media for 24 hours. The composite specimens are weighed before and after the test and the weight loss is calculated. [13,15].

3. Results and Discussion

3.1 Density and Microhardness.

The SEM image of the Co-20Al-5GNSA hybrid composites is shown in Figure.1. The SEM image is taken in secondary electron mode. It is observed that there is uniform dispersion of Co, Al, and GNSA particles in hybrid composite material. Figure.2 exhibits the microhardness and density graphs of Co-20Al-GNSA. The density of the Co-20Al-GNSA hybrid composites increases slightly with the addition of GNSA particles. The density of Co-20Al-2.5GNSA composite was 5.2 g/cm³ which increase

slightly to 5 g/cm^3 for Co-20Al-5GNSA hybrid composites. The further addition of GNSA reinforcements resulted in a decrease in the density of the Co-20Al-10GNSA hybrid composites (5 g/cm^3). The microhardness of the Co-20Al-GNSA hybrid composite increases linearly up to 5% GNSA reinforcement further addition of GNSA particles has resulted in a slight reduction in the microhardness of the composite material.

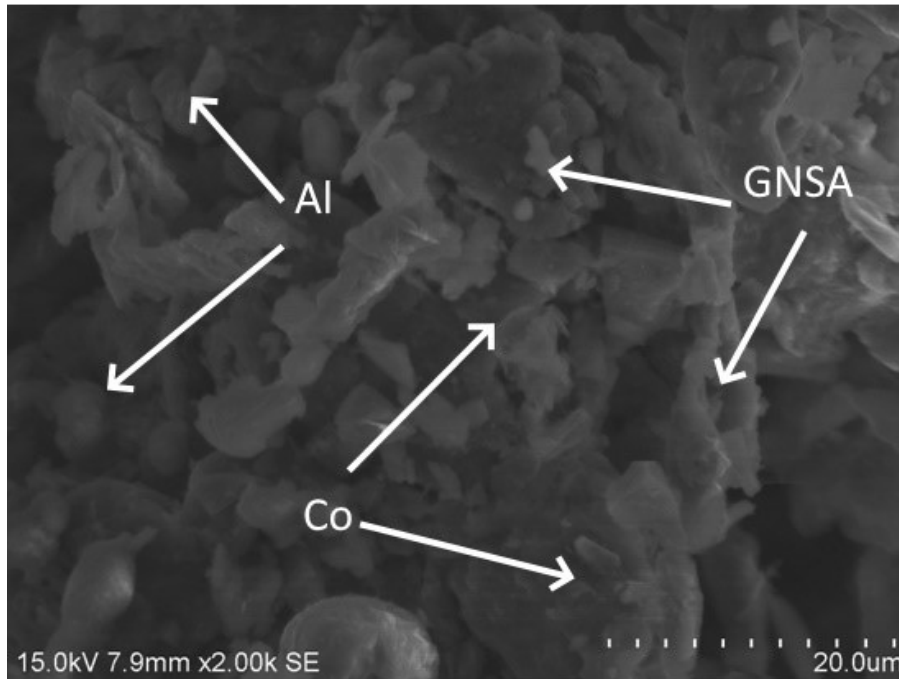


Figure.1 SEM image of Co-20Al-5GNSA hybrid composite

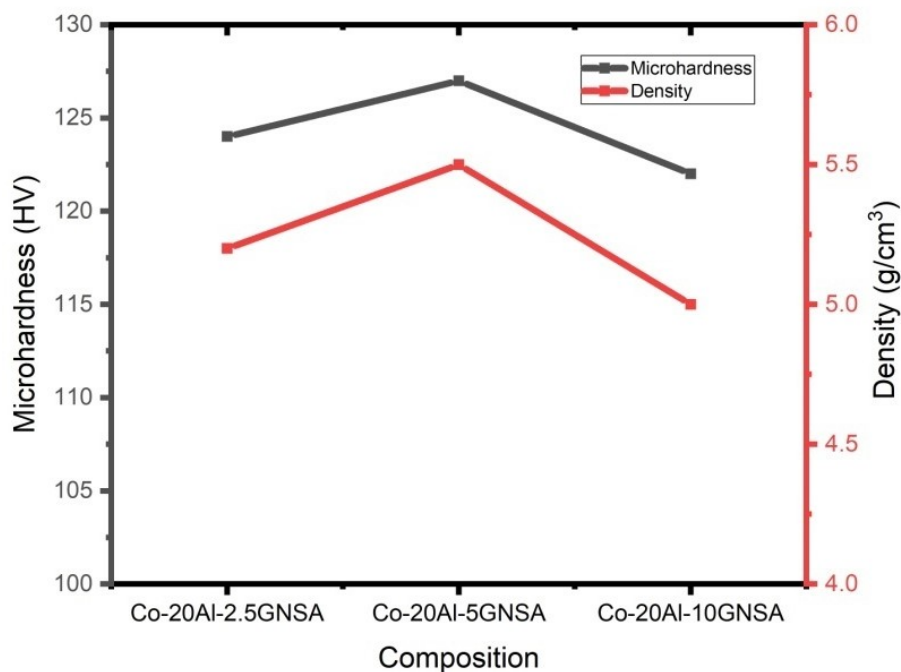


Figure.2 Microhardness and Density of Co-20Al-GNSA hybrid composites

3.2 Compressive Strength

The Compressive strength of Co-20Al-5GNSA hybrid composites is 112 MPa which is better compared to Co-20Al-2.5GNSA (110 MPa) and Co-20Al-10GNSA (108 MPa) hybrid composites. The compressive strength increased gradually with GNSA addition till it reaches 5% GNSA, whereas further addition of GNSA does not influence the compressive strength of the hybrid composites. Figure.3 represents the compressive strength of the Co-20Al-GNSA hybrid composites.

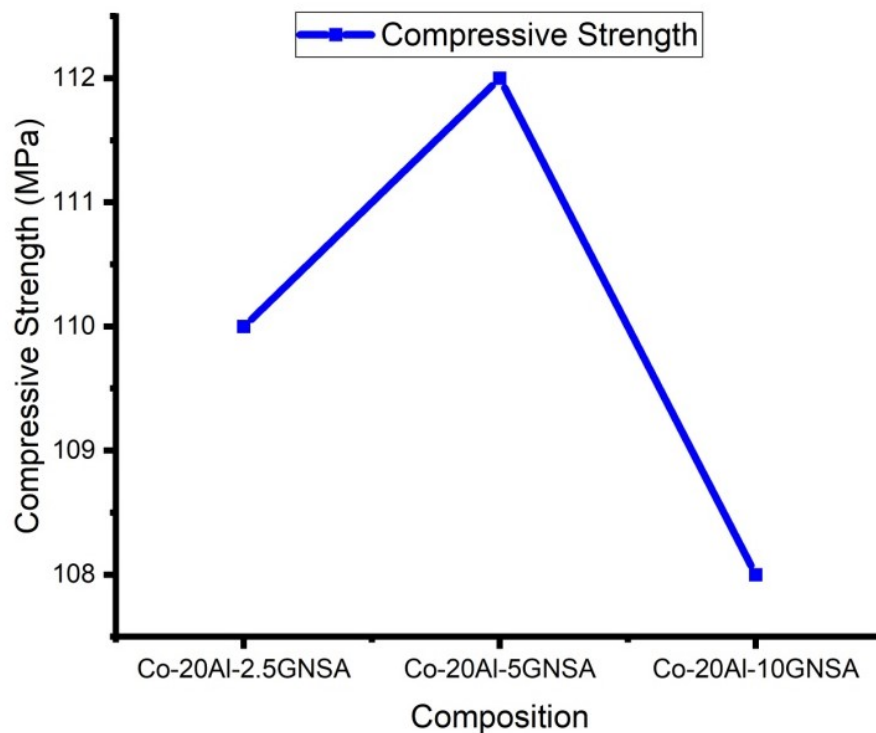


Figure.3 Compressive Strength of Co-20Al-GNSA hybrid composites

3.3 Wear and COF Analysis

The wear analysis results of Co-20Al-GNSA hybrid composites are represented in Figure.4. The wear test was carried out at a constant load of 10N, constant sliding speed of 1.5 m/s, and constant sliding distance of 1000m. The wear loss of the Co-20Al-10GNSA hybrid composites exhibited better wear resistance and COF values compared to that of other specimens. The Co-20Al-2.5GNSA has produced a COF value of 0.9 whereas the COF value of Co-20Al-10GNSA is 0.6. Hence it can be confirmed that the addition of GNSA particles has a good influence in increasing the wear resistance of the hybrid composite materials. Figure.5 represents the SEM image of Co-20Al-10GNSA hybrid composites after the wear test. From the pattern of wear track, it is evident that the major wear mechanism is abrasive wear.

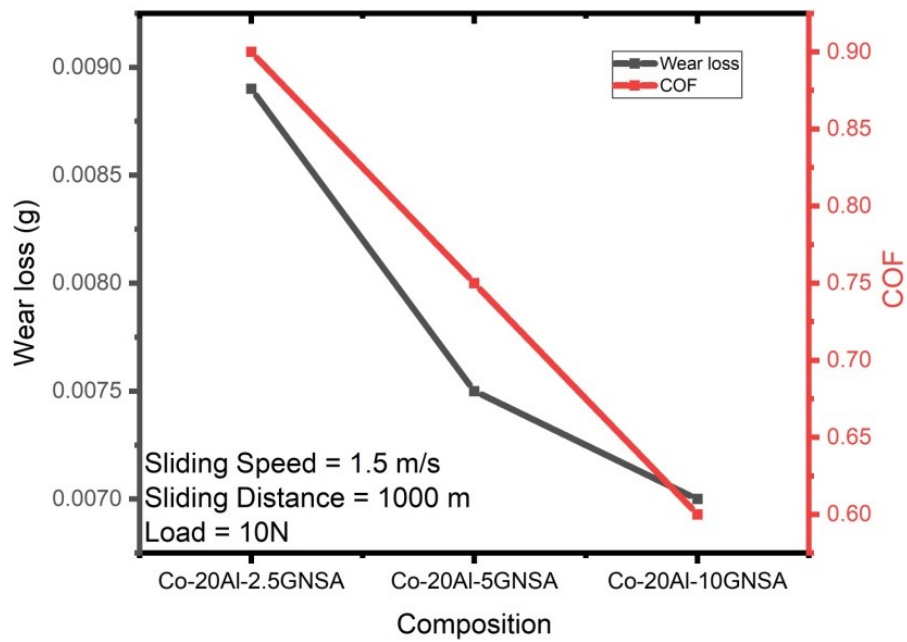


Figure. 4 Wear loss and COF of Co-20Al-GNSA hybrid composites

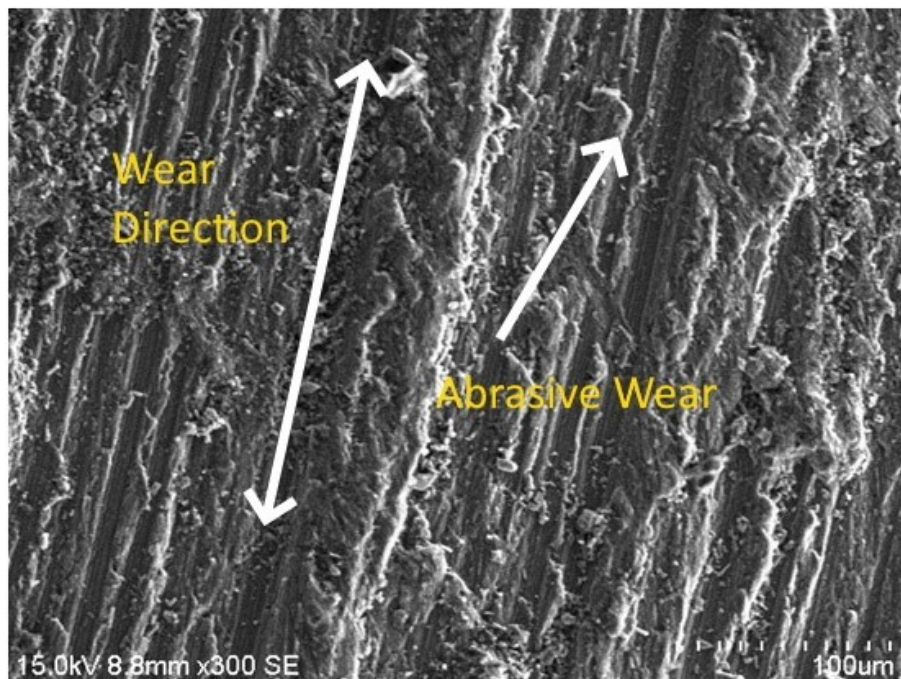


Figure .5 SEM image of Co-20Al-10GNSA hybrid composite after wear test

3.4 Corrosion Analysis

Figure .6 shows the weight-loss corrosion graphs of Co-20Al-GNSA hybrid composites at various Corrosive media such as 3% NaCl, 0.1N HCl, and 0.1N H₂SO₄. From the graph, it can be concluded that the weight loss of the composites decreases with the increase in GNSA content. The weight loss of the Co-20Al-10GNSA hybrid composite is better compared to other combinations in all kinds of corrosive media. The weight loss was maximum for 0.1 N H₂SO₄ for all samples compared to other corrosive media. The weight loss was minimum in 3% NaCl solution. The electrochemical corrosion analysis was carried out using three-electrode systems using the composite pellets as the working electrode. The potentiodynamic polarization results shows that the Co-20Al-10GNSA hybrid composites have exhibited better E_{corr} (-0.442 V) and i_{corr} values (1.5 mA/cm²) compared to that of Co-20Al-5GNSA (-0.448V & 1.7 mA/cm²) and Co-20Al-2.5 GNSA (-0.453V & 1.9 mA/cm²). It is evident that the E_{corr} values are shifted to more positive side and the i_{corr} values decreases with the increase in GNSA content which confirms the increase in corrosion resistance.

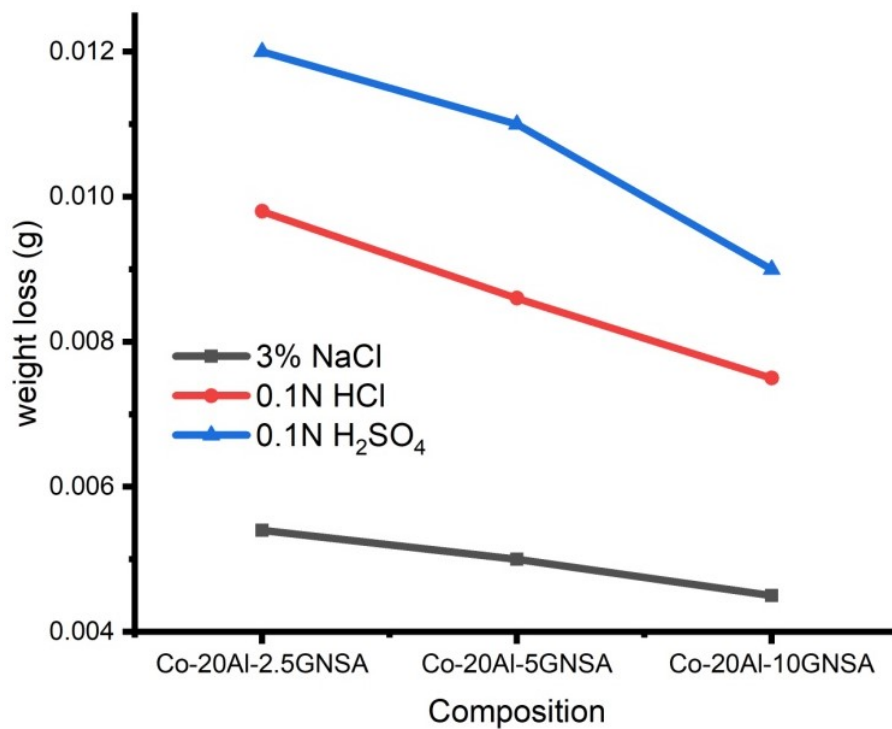


Figure.6 Weight loss corrosion results of Co-20Al-GNSA hybrid composites

4. Conclusions

The Co-20Al-GNSA hybrid composites were amalgamated by employing a high-energy ball mill.

- The Microhardness and density of the Co-20Al-5GNSA hybrid composites have improved compared to other samples.
- The compressive strength of the Co-20Al-5GNSA hybrid composite was 112 MPa and superior compared to other combinations.
- The wear analysis authenticates that Co-20Al-10GNSA hybrid composites exhibited better wear resistance and coefficient of friction.
- The potentiometric polarization analysis shows that the Co-20Al-10GNSA hybrid composites have enhanced corrosion resistance due to the existence of GNSA particles.
- The weight-loss corrosion analysis also proves that the Co-20Al-10GNSA hybrid composites have better corrosion resistance.

From the conclusion of this study, it can be accomplished that the Co-20Al-10GNSA hybrid composites have experienced a slight decrease in density, microhardness, and compressive strength but exhibits better wear and corrosion resistance. Hence it can be concluded that Co-20Al-10GNSA hybrid composites have better tribological and corrosion resistance with decent mechanical properties which may be considered for potential industrial applications.

References:

- [1] Gopinath S, Prince M and Raghav G R 2020 Enhancing the mechanical, wear and corrosion behaviour of stir casted aluminium 6061 hybrid composites through the incorporation of boron nitride and aluminium oxide particles *Mater. Res. Express* **7** 016582
- [2] Raghav G R, Balaji A N, Muthukrishnan D, Sruthi V and Sajith E 2018 An experimental investigation on wear and corrosion characteristics of Mg-Co nanocomposites *Mater. Res. Express* **5** 066523
- [3] Kumar N, Bharti A and Saxena K K 2021 A re-investigation: Effect of powder metallurgy parameters on the physical and mechanical properties of aluminium matrix composites *Mater. Today Proc.*
- [4] Albert T, Sunil J, Christopher A S, Jegan R, Prabhu P A and Selvaganesan M 2020 Preparation and characterization of aluminium-titanium carbide (Al-TiC) composite using powder metallurgy *Mater. Today Proc.*
- [5] Karthikeyan N, Krishnan B R, VembathuRajesh A and Vijayan V 2020 Experimental analysis of Al-Cu-Si metal matrix composite by powder-metallurgy process *Mater. Today Proc.* S2214785320365251
- [6] Li K C, Prior D J, Waddell J N and Swain M V 2015 Comparison of the microstructure and phase stability of as-cast, CAD/CAM and powder metallurgy manufactured Co-Cr dental alloys *Dent. Mater.* **31** e306-15
- [7] Zhou Z, Liu B, Guo W, Fu A, Duan H and Li W 2021 Corrosion behavior and mechanism of FeCrNi medium entropy alloy prepared by powder metallurgy *J. Alloys Compd.* **867** 159094

- [8] Tan L, He G, Li Y, Liu F, Nie Y and Jiang L 2018 Flow behaviors and microstructural evolutions of a novel high-Co powder metallurgy superalloy during hot working *J. Mater. Process. Technol.* **262** 221–31
- [9] Rodrigues W C, Broilo L R, Schaeffer L, Knörnschild G and Espinoza F R M 2011 Powder metallurgical processing of Co–28%Cr–6%Mo for dental implants: Physical, mechanical and electrochemical properties *Powder Technol.* **206** 233–8
- [10] Mihalcea E, Vergara-Hernández H J, Jimenez O, Olmos L, Chávez J and Arteaga D 2021 Design and characterization of Ti6Al4V/20CoCrMo–highly porous Ti6Al4V biomedical bilayer processed by powder metallurgy *Trans. Nonferrous Met. Soc. China* **31** 178–92
- [11] Venkatesh V S S and Deoghare A B 2020 Fabrication and mechanical behaviour of Al-Kaoline metal matrix composite fabricated through powder metallurgy technique *Mater. Today Proc.*
- [12] Raghav G R, Balaji A N, Selvakumar N, Muthukrishnan D and Sajith E 2019 Effect of tungsten reinforcement on mechanical, tribological and corrosion behaviour of mechanically alloyed Co-25C Cermet nanocomposites *Mater. Res. Express* **6** 1165e4
- [13] Raghav G R, kumar R A, Muthukrishnan D, Nagarajan K J, E S and V S 2020 Synthesis and mechanical characterization of Fe–BN–TiC nanocomposites *Eng. Res. Express* **2** 025036
- [14] Sudha G T, Stalin B, Ravichandran M and Balasubramanian M 2020 Mechanical Properties, Characterization and Wear Behavior of Powder Metallurgy Composites - A Review *Mater. Today Proc.* **22** 2582–96
- [15] Raghav G R, Janardhanan S, Sajith E, Chandran V and Sruthi V 2021 Mechanical and tribological performance of Al-Fe-SiC-Zr hybrid composites produced through powder metallurgy process *Mater. Res. Express* **8** 016533

Mechanical and thermal characterization of tungsten particulate reinforced epoxy polymer composites

DOI:10.36909/jer.ICMMM.12389

Jenson Joseph. E*, Kiran A.K, K. Panneerselvam

Faculty*, Department of Mechanical Engineering, PG Scholar, Production & Industrial Engineering, SCMS School of Engineering and Technology, Ernakulam, Kerala, Faculty, Production Engineering Department, National Institute of Technology, Tiruchirappalli, Tamil Nadu, India.

*Email: jenson@scmsgroup.org; Corresponding Author.

ABSTRACT

Polymer matrix composites find a wide range of industrial applications due to its unique properties like lightweight, improved strength and the properties could also be tailored to suit specific applications. In this present work, a new class of polymer matrix composites with epoxy resin as matrix and tungsten metal particles as fillers were developed. The influence of the addition of tungsten fillers on mechanical and thermal properties of the composites has been investigated. The composites are fabricated by hand lay-up method and the specimens containing tungsten particle content by 1%, 3%, 5%, 7% and 9% by weight were developed. The developed specimens were subjected to mechanical and thermal investigations. Mechanical behavior was analyzed by conducting a flexural test and hardness as per ASTM standards. Thermal behavior was analyzed by conducting Thermogravimetric analysis (TGA) and Differential Scanning Calorimetry (DSC) of the developed composites. The results show that the addition of 7 wt. % filler has a higher value of flexural strength and hardness. Further addition of particulate fillers deteriorates the flexural strength and hardness due to agglomeration of filler content in the epoxy. Analysis by TGA and DSC shows that the thermal stability of composites is improved by increasing the addition of tungsten content in the epox.



Malware visualization and detection using DenseNets

V. Anandhi¹ · P. Vinod² · Varun G. Menon¹

Received: 13 March 2021 / Accepted: 28 May 2021

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Rapid advancement in the sophistication of malware has posed a serious impact on the device connected over the Internet. Malware writing is driven by economic benefits; thus, an alarming increase in malware variants is witnessed. Recently, a large volume of malware attacks are reported on Internet of Things (IoT) networks; as these devices are exposed to insecure segments, further IoT devices reported have hardcoded credentials. To combat malware attacks on mobile devices and desktops, deep learning-based detection approaches have been attempted to detect malware variants. The existing solutions require large computational overhead and also have limited accuracy. In this paper, we visualize malware as Markov images to preserve semantic information of consecutive pixels. We further extract textures from Markov images using Gabor filter (named as Gabor images), and subsequently develop models using VGG-3 and Densely Connected Network (DenseNet). To encourage real-time malware detection and classification, we fine-tune Densely Connected Network. These models are trained and evaluated on two datasets namely Maling and BIG2015. In our experimental evaluations, we found that DenseNet identifies Maling and BIG2015 samples with accuracies of 99.94% and 98.98%, respectively. Additionally, the performance of our proposed method in classifying malware files to their respective families is superior compared to the state-of-the-art approach calibrated using prediction time, F1-score, and accuracy.

Keywords Convolutional neural networks · DenseNet · Feature maps · Malware visualization · Texture

1 Introduction

In recent years, desktops and smart devices are exposed to serious threat due to the presence of malware attacks [1]. Malware or malicious software are evolving at a faster rate [2, 3]; they are designed to disrupt, gain unauthorized access, and exfiltrate sensitive information from computer

systems. A primary motivation for developing new malware is the financial gain associated with it. Hence, it is an industry worth millions of dollars which is increasing every year. According to statistics, data breaches have increased substantially by 40% [4]. Additionally, AV-Test threat report registers [5] more than 350,000 new malware every day, and malware circulation has increased to 114,530 million in 2021; surprisingly, January 2021 alone reported the presence of 607 million malware.

Recently, malware attacks on IoT devices are increasing at an alarming rate. IoT devices have very specific functionalities such as smart healthcare (for monitoring glucose, smart pacemakers, etc), temperature monitoring particularly used in industrial control systems, smart appliances (e.g., smart refrigerator), baby monitoring systems, surveillance system using security cameras, etc. Vulnerable IoT devices of individuals and organizations are largely attacked by hackers primarily due to (a) hardcoded credentials, (b) outdated operating systems, device drivers, and (c) connection of IoT devices to an insecure network and poor web services. All these aforesaid issues transform IoT devices as a pivot to the internal network and expose them to adversary controlled servers. A widely used

✉ P. Vinod
vinod.p@cusat.ac.in

V. Anandhi
anandhi@scmsgroup.org

Varun G. Menon
varunmenon@scmsgroup.org

¹ Department of Computer Science and Engineering,
SCMS School of Engineering and Technology,
APJ Abdul Kalam Technological University,
Thiruvanthapuram, Kerala, India

² Department of Computer Applications,
Cochin University of Science and Technology, Cochin,
Kerala, India

technique for detecting malware is based on “signatures”, which is a short string of bytes that uniquely identify known malware. To detect malware, a generated signature of a file is compared with a collection of signatures stored in the database. While this approach is simple, signature detection fails to identify new malware samples, as it is highly sensitive to minor software modification, besides demand human expertise for its creation.

In general, malware detection techniques are classified as static [6] and dynamic [7] approaches. In the static method, analysis is performed on disassembled code to extract opcode sequence, API sequence, functional call graph, strings, etc. Static analysis can quickly derive program semantics but is defeated by source code obfuscation and encryption. The dynamic analysis, also known as behavioral analysis, executes the application in a controlled environment (known as a sandbox) to extract system calls or network traces, and is effective in identifying camouflaged behavior. This approach is robust but is slow and resource-intensive. Besides, malware exhibits restrained behavior on identifying analysis environments.

Generally, the performance of traditional machine learning algorithms depends on robust feature extraction [8, 9] methods. This resulted in the popularity of machine learning in the security domain, especially malware detection [8, 10, 11]. However, these approaches are not effective in detecting new variants or unknown malwares [12]. Recent research has demonstrated the popularity of deep learning techniques in numerous applications such as object detection, image segmentation, medical image analysis, and fraud detection. Hazra et al. [13] exhaustively reviewed diverse deep learning methods and prioritized Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory, and Deep Belief Network models as the most promising architectures based on its acceptance in modern applications. Furthermore, deep learning techniques, in particular Convolutional Neural Network (CNN), gained importance in detecting malicious software [14]. Consequently, many intelligent anti-malware solutions with deep learning methods [15–18] have been reported.

A vast majority of malware samples have structural dissimilarity having certain functionality in common. The collection of such malicious programs is known as “family” [10]. In this scenario, the detection of different types of malware is a major concern to both researchers and the anti-virus industry. It is reported that less than 2% code difference exists in newly generated variants. To minimize the domain knowledge and feature extraction cost, researchers have addressed malware detection problems as image visualization techniques, largely due to the presence of visual similarity among variants of the same family. In this approach, a malicious binary is transformed into a two-dimensional image [10], and these images are presented

to the machine learning or deep learning classifiers for extracting relevant features for classification. In this way, malware scanners will be agnostic to the operating systems and file types. However, simply representing malware binaries as 2D images is disregarded as they lack semantic information. Additionally, training algorithms on the entire pixels of an image is computationally expensive. Thus, such malware scanners will be impractical to deploy on the devices.

To address the above-stated limitations, we created a deep learning-based malware classifier capable of detecting and classifying malicious executables in real-time. Due to resource constraint in IoT devices, we perform the analysis of malicious samples in the cloud environment. The system used by the end-user periodically transfers executable to our analysis environment deployed on the cloud, which performs data preprocessing and prediction of new samples using deep learning models. Our approach maps executables as grayscale and color images and textures are extracted using Gabor filter, assuming a prevalence of code reuse in new variants. We call such images Gabor images. Subsequently, we create classification models trained on Gabor images using a deep convolutional network and DenseNet. Finally, we prove through comprehensive experiments that DenseNet can accurately detect malicious executables and classify the samples into their respective family in real-time.

In summary, the main contributions of this paper are as follows:

- We develop a malware detection and classification system using different types of visual features (grayscale, RGB, and Markov image). This is achieved by training fine-tuned Convolutional Neural Network (specifically VGG3) and Densenet over two benchmark malware datasets, i.e., Maling and BIG 2015.
- We conduct comprehensive experiments on image texture and show significant improvement in the performance of deep learning models comparing standard visual features (gray and RGB images).
- We compare our proposed solution with the state-of-the-art approaches using the identical datasets for a fair comparison. Furthermore, through intensive experiments, we show that our approach surpasses the previous studies both in execution speed and classification results. In addition, we created adversarial samples by injecting additive noise into files. Finally, we observed a marginal drop in the performance of DenseNet. This proves the efficacy of the trained model for detecting new samples.

Section 2 gives a brief introduction to the related work. Our contributions right from the preprocessing to deep learning classification methods are described in Section 3.

Section 4 explains the evaluation metrics used in this paper. The experimental results and discussions are shown in Section 5. Section 6 concludes the paper with directions for future work.

2 Related work

In this section, we present the related research papers based on feature analysis and malicious code visualization.

2.1 Feature analysis

The main approaches of analyzing malware based on feature analysis are static and dynamic code analyses. In static analysis, the code is disassembled and examined for malicious patterns. Dynamic analysis is a behavior-based procedure where the code is executed in a virtual environment and the execution trace is analyzed. The static analysis offers the complete analysis procedure but suffers from code obfuscation. Dynamic analysis is more efficient but the main disadvantage is a time-consuming process and resource-consuming too. Some malicious apps might not be observed because the environment is only a virtual one which does not satisfy all conditions of a real one.

The static analysis method analyzes binary values or hex values and extracts some features from the files. Machine learning methods can be applied to the static features obtained. It is observed that this method can achieve better accuracy. Previous studies of static methods use permissions to verify and check application risk, but these methods are difficult to guarantee high accuracy [6, 19]. A method based on static analysis is proposed in [6] for malware detection in Android devices which is performed by extracting specific feature information like serial number, and based on this classify malicious apps of similar apps. Nezhadkamali et al. [20] proposed a permission-based algorithm that uses Intents and API sequences. The method modifies the obtained values using the same features maps. In this way, they were able to obtain accuracy to 98.6%.

Vasan et al. [21] proposed a data mining method that detects malware variants by static methods. The analysis introduced extraction of features and efficiently classifies malware family which solves the data imbalance problem with the help of the augmentation technique.

The dynamic analysis method executes the files in a virtual environment. Shezan et al. [22] proposed a detailed study on awareness to the users about common vulnerabilities present in the OS and the quality tools for testing the vulnerabilities.

Han et al. [7] proposed a method based on correlation with semantics and a combination of static and dynamic API functions and construct an overall feature map vector space

for malware detection. In [23], a probability matrix-based classification system and machine learning are developed for static and dynamic analyses using CNN with spatial pyramid pooling techniques, where the accuracy of 98.82% is achieved. Yoo et al. [24] proposed a hybrid model consisting of a Random forest classifier and deep learning models with 12 hidden layers to identify malware and benign files, and obtained a detection rate of 85.1%. In short, to detect malicious code a behavior model is established. These behavior models achieved better detection results but developed unreliable results. Moreover, dynamic analysis consumes more processing time as the computational overhead is more, which leads to less accuracy when large datasets are involved.

2.2 Malicious code visualization

Nowadays, many tools can visualize the available data. Several studies propose malware classification using visualization techniques. The fundamental idea employed in these approaches is to dissect the patterns of malicious apps and new malware variants. It also identifies the changes made on benign software due to viruses. The analysis also finds the relation between different malware families and can identify the unknown malware family too. Recent studies point out that the patterns of benign and malicious apps are different [25]. Experiments performed using CNN in this paper report an accuracy of 90.67%. Furthermore, this paper concludes the existence of relevant features for classifying samples into their respective family.

Nataraj et al. [10] proposed a novel method for malware detection, where the binary executables packed in families are converted into grayscale images of fixed size based on the technique of texture analysis. Then, the features are used for the detection of malware and the efficiency was better in less time.

Han et al. [26] proposed a dynamic analysis visualization procedure to obtain RGB-colored pixels using the opcode values from malware files and the similarities are calculated and the procedure applies to packed malware samples too.

Zhong and Gu [12] proposed a malware detection method on deep learning by using multiple levels, where each model does not work on the entire dataset, but it works on part of the dataset for a group of malware families so that each model can learn a particular data distribution. Finally, all the models together decide hierarchically and make a final decision on the data.

Due to the rapid growth of malware variants, a classification method based on Markov images and deep learning [27] was introduced using byte-level. It converts the binaries into Markov images on the two standard malware datasets, the Microsoft dataset and the Drebin

dataset and the average accuracy rates obtained are 99.264% and 97.364% respectively.

Deep learning has been a solution in many malware classification applications. In [28], a deep learning model classifier was developed for CNN-based architecture to classify malicious samples which achieved accuracies of 98.52% and 99.97% on the Microsoft datasets and Maling datasets respectively. Roseline et al. [29] proposed a system that does not require backpropagation or hyperparameter tuning, thereby reducing model complexity. The model has detection rates of 97.2% and 98.65% for BIG 2015 and Maling datasets respectively. Their model deficits in identifying unknown samples of untrained families.

Gibert et al. [30] developed a deep learning system that learns visual features of Maling and BIG2015 malicious executables to classify them into families. To visualize, each malware sample was converted to a grayscale image. A fine-tuned CNN model was developed which classified malicious files of both the datasets with accuracies of 98.48% and 97.49% respectively. Executables of Microsoft BIG2015 were transformed to opcodes and finally represented as images [31]. Then, GoogleNet model and ResNet model were applied to detect malicious files. The ResNet attained 88.36% precision, whereas 74.5% accuracy was obtained on GoogleNet.

Hamad Naeem et al. [32] developed a hybrid approach on color images to detect IoT malware using two datasets. A deep CNN model was selected to extract visual features. The detection accuracy was reported on two datasets as 97.81% and 98.47% respectively. Recently, Turker et al. [33] proposed a cognitive and intelligent anti-malware model for classifying samples of the Maling dataset. They created hybrid features, extracted through LBP-SVD-LPNet. Subsequently, Principal Component Analysis (PCA) was applied to obtain relevant attributes, and finally, a classification model was developed using Linear Discriminant Analysis (LDA). This method achieved 87.62% F1 score and 88.08% detection rate. Çayır et al. [34] proposed Random Capsule Network (RCNF) reducing the variance of different CapsNet. They conducted the experiments on a public benchmark dataset, i.e., Maling and BIG2015. The authors reported that the RCNF achieves F1 scores of 0.966 and 0.982 on Maling and BIG2015.

Sun and Qian [35] proposed a method to combine the static analysis with recurrent neural networks and CNN. By using this technique, the accuracy obtained was over 92%. By increasing the training dataset, the accuracy reached 99.5%. Feng et al. [36] proposed an Android malware detection system in a real-time environment on mobile devices and found that memory usage was 60 MB and the execution time was 0.46 s. Detailed analyses on accuracy and performance on various mobile devices were performed

and accuracy obtained was 96.75% on all devices with a response time of fewer than 3 s.

To detect and classify malware samples in real-time, we propose the implementation of DenseNet, trained on visual artifacts of the malicious sample. Another important aspect of our proposed approach is the use of texture-based images for representing executables. This reduces computational overhead by only allowing the classifier to remember prominent patches of the image which acts as its global feature.

3 Proposed method

Malware writers usually change a part of the previous codes available to produce new malware [37] variants. If we try to represent malware as an image, these changes can be detected by robust malware visualization techniques. With this idea, we visualize malware binary files as grayscale images, RGB images, and Markov images. Later, we perform texture analysis to determine specific components around the region of interest.

Figure 1 shows the overall architecture. In this system, the benign and malware files are represented as fixed size images. The benign executables were collected from various sources which include laptops and the Internet [38–41]. Then, we use VirusTotal to check if the collected executables are malware free. In addition, public benchmarks, i.e., Maling [10] and BIG2015 [42], are considered as malware datasets. We choose to study on these samples as (a) both the datasets have class imbalance, this will help us evaluate the robustness of the proposed model, and (b) the samples in the aforementioned dataset are exhaustively evaluated in a large collection of prior work [12, 27, 30, 37, 43]. All binaries including malware and goodware were converted to images. Subsequently, these image files were supplied to DenseNet which learns appropriate features necessary for the task of classification.

3.1 Preprocessing

Our system transforms executables to grayscale, RGB, and Markov images. The steps for converting a given binary into an image are explained in the subsequent paragraphs.

3.1.1 Grayscale images

The given hexadecimal values of executables are converted into binary. We read bytes and express it as gray values in the range of 0–255, where 0 indicates black and 255 represent white. These decimal values are organized as a two-dimensional matrix of size 256×256 .

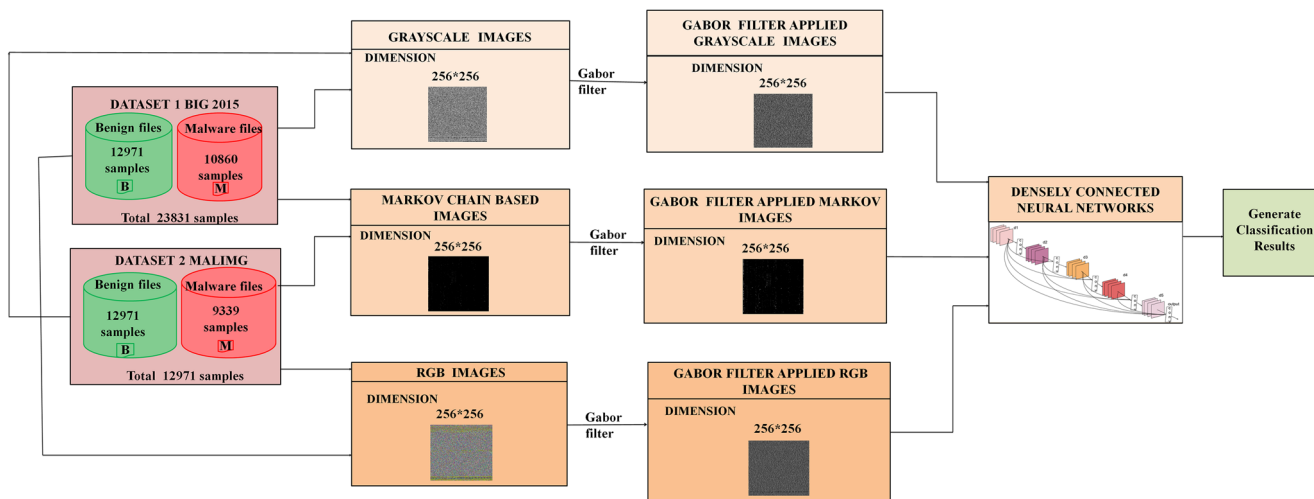


Fig. 1 Overall architecture of malware detection method

3.1.2 RGB images

In this process of visualization, we represent malware and legitimate binaries as RGB channels. To accomplish this task, we create a color map, a two-dimensional data structure having 256 rows and 256 columns, consisting of pixel values. For converting an executable to an RGB image, we read two bytes at a time, the first byte is used as row index, and the second byte as column index of the color map. Finally, the element at the intersection of the row and column indices is chosen as a pixel in the red, green, and blue plane.

3.1.3 Markov images

Figure 2 shows the steps for converting binaries into Markov images. Generally, for creating visual features, one byte is read and mapped to an equivalent pixel [10]. The influence of truncating and padding bytes to a file on classifier accuracy is not precisely addressed in prior studies. To circumvent the above-stated issue, Markov images are developed, where the bytes are considered as a stochastic process. This means that probability of byte b_i is related to byte b_{i-1} ; thus, the Markov chain is represented as:

$$P(b_{i+1}|b_0 \cdots b_i) = P(b_{i+1}|b_i)$$

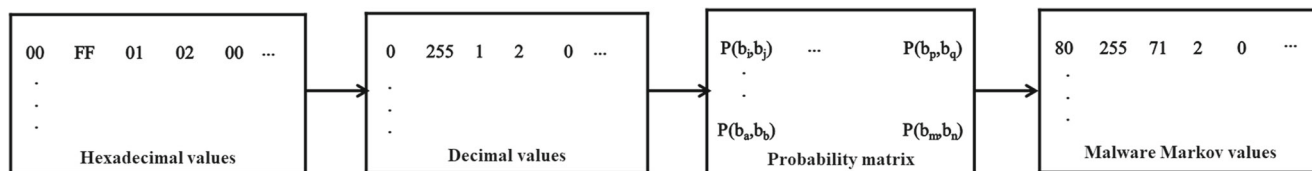


Fig. 2 Generation of Markov images

The steps to generate Markov images are summarized below:

- **Step 1:** First a frequency matrix is created which records the occurrence of consecutive pixel pairs denoted by $f(x, y)$, for bytes x and y .
- **Step 2:** Using the frequency matrix, we create a Markov probability matrix where the elements are transfer probabilities of byte x and y . Probabilities in the Markov probability matrix (see (1)) are populated by taking each element in the frequency matrix and dividing by the corresponding row sum.
- **Step 3:** Finally, we map the probability to pixels. Here, each element in the probability matrix is multiplied by 255 and divided by the maximum probability to generate the Markov image of size 256×256 .

$$P(x, y) = P(y|x) = \frac{f(x, y)}{\sum_{n=0}^{255} f(x, y)}. \tag{1}$$

3.1.4 Gabor images

Gabor filter is popularly used in diverse computer vision applications mainly in object detection, feature extraction, and image segmentation. These filters mimic the mammalian visual cortex to recognize textures in an image. A 2D-Gabor filter derives features in both the spatial and

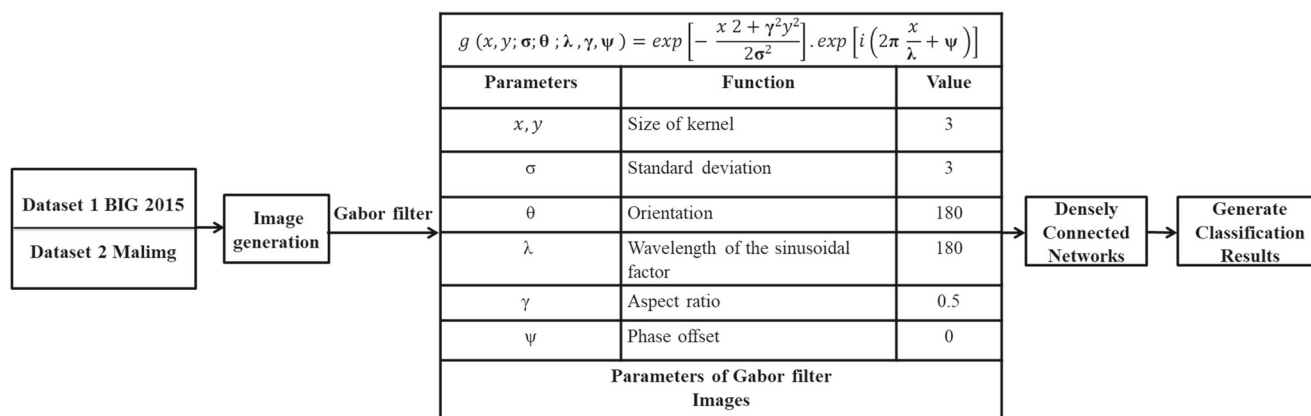


Fig. 3 Parameters of Gabor filter applied to images

frequency domains. A Gabor function is the product of sinusoidal and Gaussian functions in the time domain, and convolution of transforms of these functions in the frequency domain (refer to Fig. 3). Gabor filter analyzes whether there is a particular frequency of information in that particular direction around the region of analysis. We can obtain a bank of Gabor filters by varying the frequencies and orientations in the Gabor function. The parameters at which we obtained better results are also shown in Fig. 3. An input image is convoluted through a set of Gabor kernels to obtain image texture. The figure briefly presents the steps for generating Gabor images. Later, Gabor images are fed to CNN and DenseNet for classification and prediction.

3.2 Feature extraction and learning

Recently, researchers have attempted using deep learning techniques for classifying and detecting malicious files. As convolutional neural networks (CNN) and densely connected networks (DenseNet) have proven to be very efficient for computer vision problems [44, 45], we thus adopt aforesaid architectures for malware classification using visual features. In the following subsections, we introduce CNN and DenseNet.

3.2.1 Convolutional neural networks

Convolutional neural networks have always become the popular machine learning algorithm [14] for image classification and object detection. CNNs have two important operations: (a) convolution and (b) pool. In the forward pass, the convolutional layer utilizes a bank of filters, each of which convolves across the width and height of the executables image to output feature maps. Next, the max pooling layer progressively downsamples the feature map. Max pooling takes the maximum value by sliding 2×2 window over the feature map. The combination of convolution

followed by max-pool is a new image that is further presented to subsequent layers of CNN. Finally, the output of the max-pool is flattened and given as input to the fully connected network which is further connected to the output layer. The class of malware file is given to the output through softmax activation function for malware classification and sigmoid activation function for the malware detection task. During the training phase, the values of the parameters are updated continuously until the learning is complete. In our study, we use the deep convolutional network, in particular VGG3 [46]. VGG3 learns representative attributes of a sample without the need for manual feature extraction by performing multiple nonlinear transformation to produce abstract representation for each executable image.

CNN has found more importance in medical applications like chest disease detection [47, 48]. Choudhary and Hazra employed the VGG16 model to train the medical images and the results obtained are 98% of training accuracy and 97% of test accuracy.

3.2.2 Densely connected networks

The motivation for selecting DenseNet is circumventing the vanishing gradient problem. For deeper networks, gradients are not propagated back to the previous layers or the initial layers. The gradient shrinks for a deeper network due to which the initial layers do not learn. If there is no significant change in the gradient, it is understood that the weights are not updated or in effect no learning takes place. Hence to resolve the above-stated issue we have considered DenseNets. DenseNets have interconnection paths where features are combined by concatenation. The main highlights of the DenseNet architecture are feature reuse, vanishing gradient problem, feature propagation, and less parameter count [44]. These interconnections form deep and dense paths that learn better than other networks.

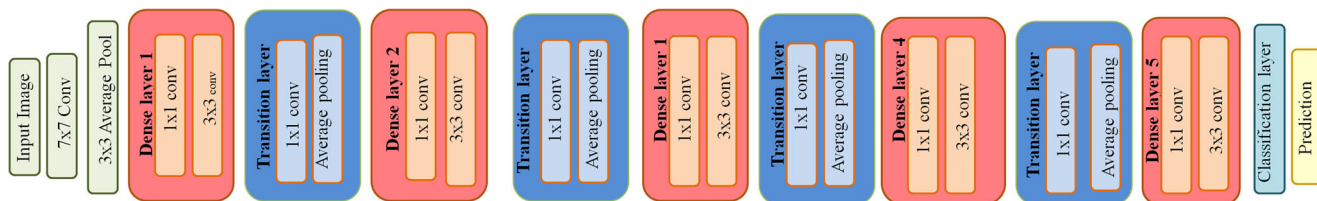


Fig. 4 Detail architecture of DenseNet

The DenseNet comprises dense blocks, composite function, and transition layer. First is the dense block where each block comprises n dense layers. These dense layers are connected such that each layer receives information from all preceding layers and passes its information to all subsequent layers. Thus, each layer receives collective information from all the subsequent layers. The width and height of the feature maps stay the same in a particular block. Each layer consists of two convolutional operations: 1×1 convolution (CONV) for extracting features and 3×3 CONV for bringing down the feature count. The composite function consists of batch normalization, Rectified Linear Unit (RELU), and 3×3 convolution layer. This function concatenates the output of the preceding layers to the next layer.

Finally, the transition layer consists of 1×1 CONV operation which reduces the parameter count, followed by the 2×2 average pool layer, responsible for downsampling the dimension feature space. Figure 4 illustrates the detailed architecture of DenseNets. The whole idea of DenseNet is to reuse features. Additionally, maximize the flow of information between dense layers. As the model requires only a fewer number of layers, they are easy to train and a lesser number of parameters need to be learned. In our study, we used a specific variant of DenseNet, i.e., DenseNet201, consisting of 201 layers deep. The motivation of using DenseNet201 was drawn due to improved performance as reported in [44]. Additionally, DenseNet has a strong gradient flow. The error signals are transferred to earlier layers compared to the conventional deep neural network. Thus, the earlier layers timely administer errors without waiting longer for gradients, computed from previous layers.

Table 1 Number of trainable parameters in the proposed DenseNet model

Hyperparameters used in the model	Value
No. of layers in DenseNet	201
Learning rate	0.001/0.01
No. of epochs	30/50
Batch size	32/64

In our experiments, we have used DenseNet architecture with the number of trainable parameters as shown in Table 1.

4 Evaluation metrics

The evaluation metrics used in our work for the classification and malware detection are Precision, Recall, F1 measure, and Accuracy. These metrics are calculated using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP indicates the number of malware samples correctly identified as malware. TN indicates the number of benign samples accurately identified as legitimate. FP is the number of benign samples misclassified as malware. FN indicates the number of wrongly malware images. The evaluation metrics are defined using (2) through (5).

$$Precision(P) = \frac{TP}{TP + FP}, \tag{2}$$

$$Recall(R) = \frac{TP}{TP + FN}, \tag{3}$$

$$F1measure(F) = 2 * \frac{R * P}{R + P}, \tag{4}$$

$$Accuracy(A) = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5}$$

The performance of DenseNet on different types of images is compared and shown in Figs. 5, 6, 7, and 8. An ideal malware detection model should have a high value of F1 measure. For the Gabor filter applied to Markov images, F1 measures obtained are 98.73% and 99.94% for BIG 2015 and Maling datasets, respectively.

5 Results and discussions

Implementation and setup The hardware requirements for conducting the experiments are Intel(R) Xeon(R) CPU @ 2.20GHz*2 processor with 12 GB RAM capacity and 1 TB hard drive. Experiments were conducted on Ubuntu 18.04 platform. Instead of normal CPU computing, the use of NVIDIA Tesla P100 16 GB graphics card increases

Fig. 5 Performance measures of DenseNet without applying Gabor filter, Dataset Benign + BIG 2015

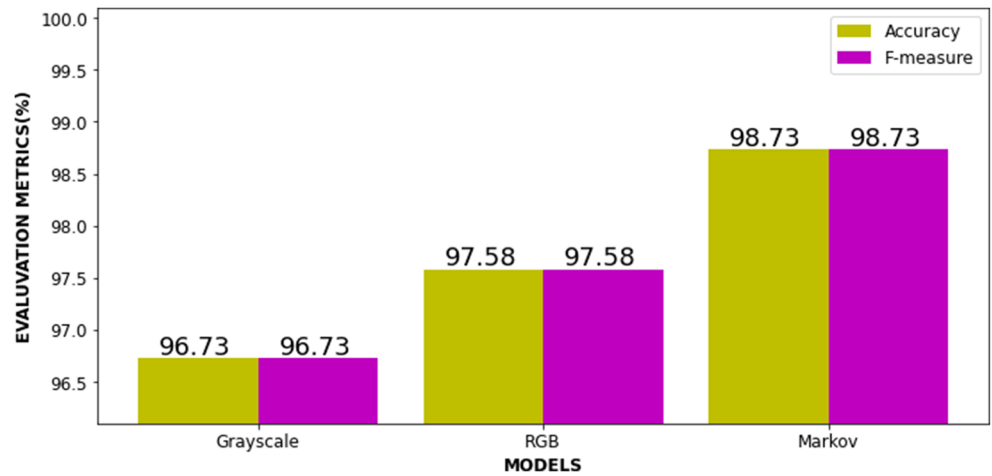


Fig. 6 Performance measures of DenseNet without applying Gabor filter, Dataset Benign + Maling

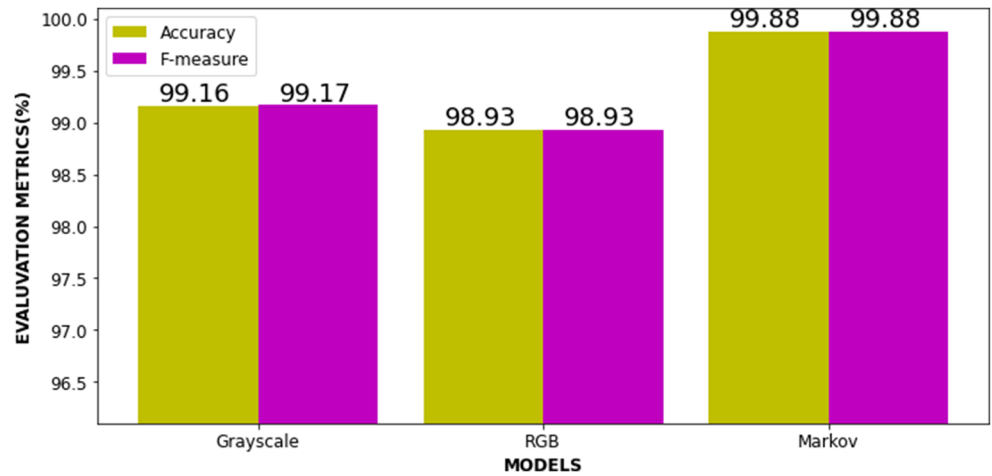


Fig. 7 Performance measures of DenseNet using Gabor filter applied on 256x256 images, Dataset Benign + BIG 2015

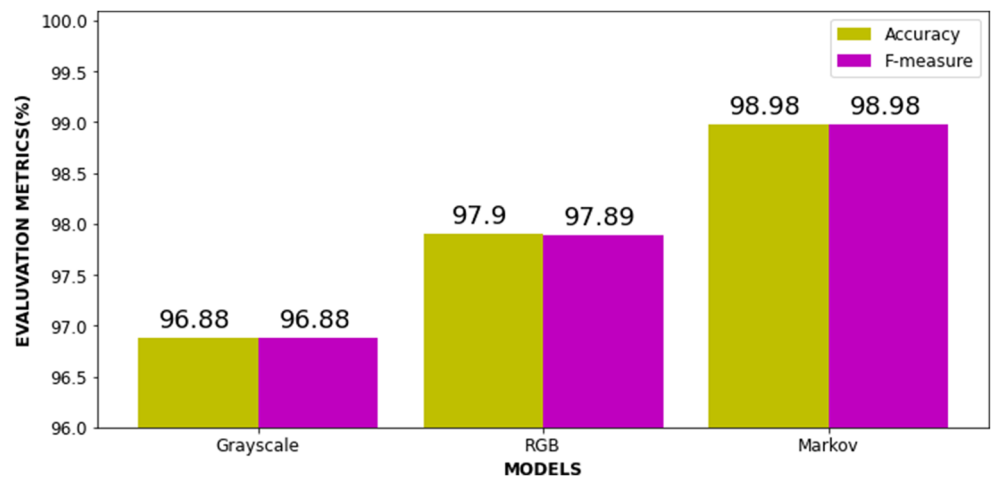
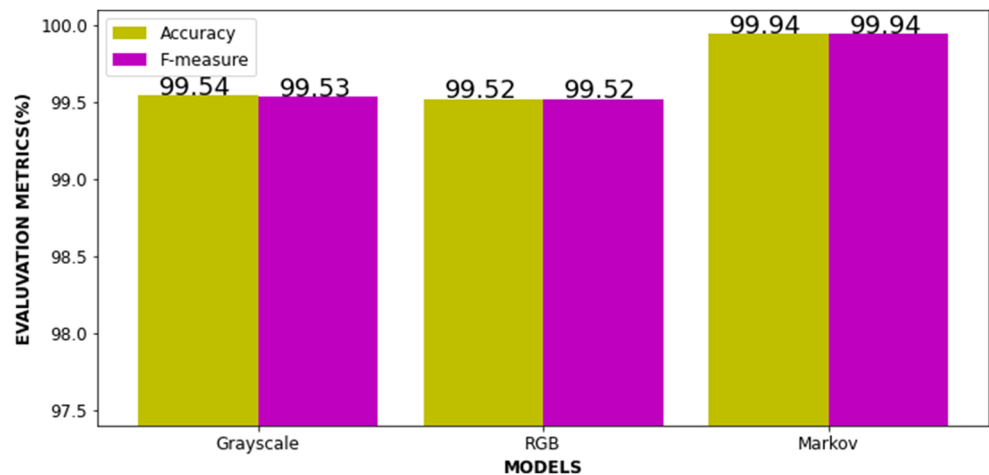


Fig. 8 Performance measures of DenseNet using Gabor filter applied on 256×256 images, Gabor filter applied images, Dataset Benign + Maling



the computational speed. The software implementation was done using Python 3.7.6 programming language, with packages included such as TensorFlow 2.2.0 and Keras 2.3.1 with Scikit-learn machine-learning library and Matplotlib for visualizing results.

Dataset The experiments are performed on two public benchmark malware datasets, BIG 2015 [42] and Maling [10]. Maling dataset has 9339 samples grouped in 25 families; all binaries are converted to grayscale images and published for use. BIG2015 dataset has been released as Kaggle competition; this dataset contains 10868 ASM and byte files corresponding to 9 different malware families. The benign dataset samples constituted of executable collected from [38–41]; furthermore, each executable was checked for malicious files by uploading the file to VirusTotal.

For conducting experiments, we created two datasets. Dataset 1 comprises 23,831 samples where 12,971 samples were benign and 10,868 BIG 2015 samples. Out of the 12,971 benign samples, 9080 samples were used as train set and 3891 samples were used as a test set. From the BIG 2015 malware dataset, 7602 samples constituted the train set and 3258 samples were reserved as the test set. In addition, we created Dataset 2 having the same number of the benign sample as in Dataset 1, but has 9339 Maling images. From the Maling malware dataset, 6537 samples were used as a train set and the remaining 2802 files were used for prediction. All images were resized to 256×256 for CNN and DenseNet. We conduct the following experiments:

- **Experiment-1** Effect of image representations on classification performance.
- **Experiment-2** Performance of DenseNet in detecting executables.
- **Experiment-3** Comparison of VGG3 and DenseNet in predicting malicious executables.
- **Experiment-4** Classification of malware samples into their respective families (i.e., multi-class classification).

- **Experiment-5** Comparative analysis of proposed model and state-of-the-art malware detection/classification methods.
- **Experiment-6** Performance evaluation of DenseNet model on evasion attack.

5.1 Effect of image representations on classification performance

In this section, we compare classifier results obtained while inputting different types of images. Table 2 shows the results of the prediction obtained for two datasets. An F1-measure of 98.73% is achieved for BIG2015 samples, and 99.88% F1-measure is obtained for Maling samples during prediction. The best classifier performance is obtained with Markov images, followed by RGB, and inadequate results are attained for grayscale images. Additionally, improved performance is obtained if textures are extracted from images and presented during the training phase. We notice an increase in F1-measure for BIG2015 Gabor images (grayscale, RGB, and Markov); improvements in F1-measure attained are 96.88%, 97.81%, and 98.98% respectively. Similarly, F1-measures achieved on Maling dataset are 99.45%, 99.52%, and 99.94% respectively.

In all cases, we can visualize that the classifier performance is limited when trained on grayscale images. Furthermore, the results of grayscale BIG2015 images are less compared to Maling. Especially in the case of BIG2015, executables were disassembled using IDAPro, and such tools generate output with a fixed structure. In particular, we observed a large volume of black pixels in the grayscale images. This would cause the max pool layer to always output identical values, and such dark regions do not contribute meaningful information during the learning phase. Nevertheless, we also witness that deeper network structures especially VGG3 tend to output moderate results and tend to capture the localized association between the

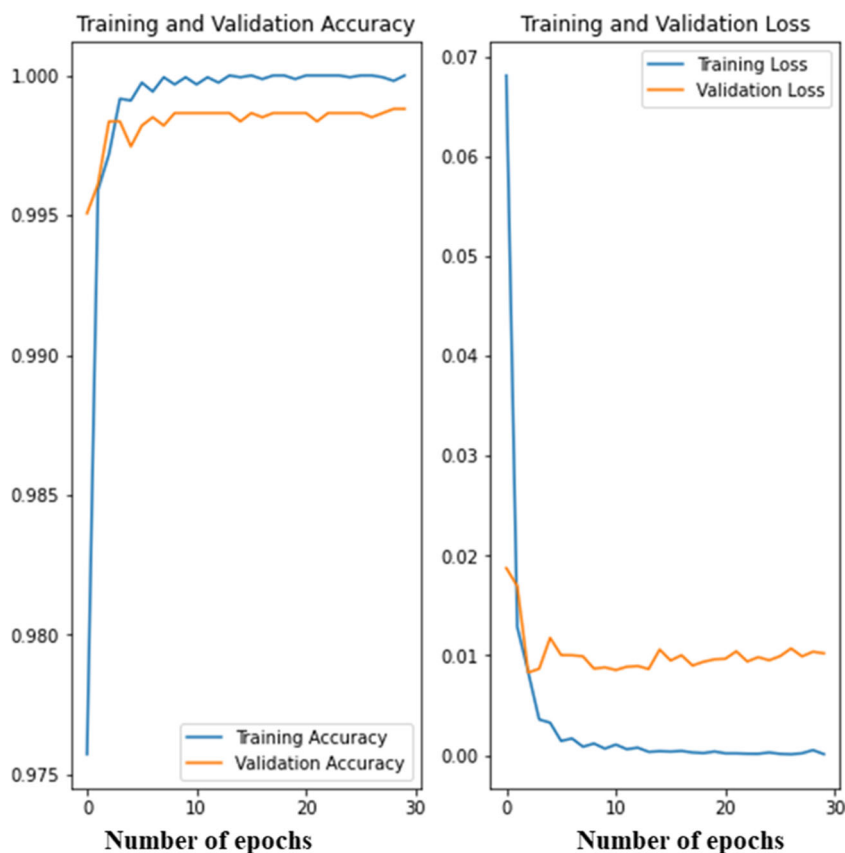
Table 2 Performance evaluation of convolution neural network and DenseNets without and with using Gabor filters

Dataset	Image	Architecture	Without Gabor filter		With Gabor filter	
			A (%)	F (%)	A (%)	F (%)
BIG 2015	Grayscale	VGG3 Dropout	96.37	96.38	96.39	96.39
		DenseNet	96.73	96.73	96.88	96.88
		VGG3 Baseline	97.38	97.38	97.89	97.89
	RGB	DenseNet	97.58	97.58	97.90	97.89
		VGG3 Dropout with Batch normalization	97.96	98.81	98.81	98.81
		DenseNet	98.73	98.73	98.98	98.98
	Markov	VGG3 Dropout with Batch normalization	98.72	98.72	99.45	99.45
		DenseNet	99.16	99.17	99.54	99.17
		VGG3 Baseline	97.83	97.84	99.21	99.21
		DenseNet	98.93	98.93	99.52	99.52
Malimg	Markov	VGG3 Dropout with Batch normalization	99.16	99.16	99.82	99.82
		DenseNet	99.88	99.88	99.94	99.94

pixels. However, superior outcomes are attained in the case of Markov images. Primarily, extraction of texture improves detection, as such images carry minute details such as dots, lines, and boundaries separating different regions in the image. Consequently, the best results are achieved with Markov-Gabor filter images. In addition, we

observe moderate accuracy with classifiers trained on RGB images, although RGB images enhance changes in sub-region relatively better compared to the grayscale images. Hence, such differences in regions are easily captured by the deep classifiers during the training phase, and eventually, distinguish malware from their families.

Fig. 9 Comparison of training and validation loss and accuracy with validation and testing samples



5.2 Performance of DenseNet in detecting executables

From Figs. 5, 6, 7, and 8, it is clear that F1-scores of 99.94% and 98.88% are attained for Maling and BIG2015 respectively. Comparing all figures, it can be seen that highest F1-score and accuracy are obtained if Gabor filter applied Markov images were used to train the DenseNet. Figure 9 depicts the accuracies on the train and validation sets. Likewise, the validation loss on the dataset is reported in the same figure. It is evident that the accuracy improves beyond 10 epochs and converges between 25 and 30 epochs. An identical trend is obtained while we evaluate the loss. In particular, beyond 25 epochs convergence in the loss can be witnessed. Furthermore, in Table 3, we compare time for predicting new samples along with hyperparameters for different DenseNet architectures. We can visualize from Table 3 that the average execution time for predicting a new sample in both datasets is in the range of 0.004 to 0.006 ms which matches real-time scanning of files by antivirus software.

From all experiments, we see that DenseNet obtains a very high F1-score and accuracy particularly due to two fundamental reasons: (a) deeper architecture and (b) interconnections with the previous layers. The deeper network advocates each layer to extract executable image pattern which is disparate from other layers. Besides, increased inter-connectivity of layers allows the re-use of feature maps of previous layers and creates a new image representation that carries information from all predecessor layers along with the current layer. This resolves the problems induced with redundant layers which are typical in deeper architectures.

5.3 Comparison of VGG3 and DenseNet in predicting malicious executables

Table 2 compares the performance of VGG3 and DenseNet on different types of images. Here, we see that for BIG 2015 dataset represented in the form of Gabor images; VGG3 shows a maximum F1-score of 98.81%. While for the identical images, DenseNet attains F1-score of 98.98% which is 0.17% higher than VGG3. Similar trends in the results were observed for Maling samples. It can be observed that DenseNet accuracy and F1-score for Gabor images is 99.94%, while for VGG3 an F1-score of 99.82% was obtained. Even though we found marginal improvement in the DenseNet performance, the prediction time for BIG 2015 with this model is 0.006ms, and 0.004ms for Maling files. In addition, we also found a reduction in the number of parameters for implementing DenseNet architecture comparing the VGG3 models.

Table 3 Network topology and performance metrics for image dimension 256 × 256 by applying Gabor filter with detection time

Image type	Maling			
	Benign + BIG 2015	Hyper-parameters	Prediction time (ms)	Network topology
Grayscale	DenseNet201+	lr=0.01, epochs=50, batch_size=64	Total time= 27.107, Avg time= 0.004	DenseNet201+
	Dense(1024)+			Dense(1024)+
	Dropout(0.1)+			Dense(2048)+
RGB	Dense(2048)+	lr=0.01, epochs=50, batch_size=64	Total time= 39.871, Avg time= 0.006	Dense(2048)+
	Dropout(0.1)+			Dense(4096)
	Dense(4096)+			Dense(4096)
Markov	Dropout(0.1)	lr=0.01, epochs=50, batch_size=64	Total time= 40.857, Avg time= 0.006	DenseNet201+
	Dense(1024)+			Dense(1024)+
	Dropout(0.1)+			Dense(2048)+
Markov	Dense(2048)+	lr=0.001, epochs=30, batch_size=32	Total time= 25.581, Avg time= 0.004	Dense(2048)+
	Dropout(0.1)+			Dense(4096)
	Dense(4096)+			Dense(4096)

Summary From experiments 1, 2, and 3, we conclude that different deep learning models trained with Gabor Markov images accurately represented malware and benign executables comparing other types of images (grayscale and RGB). Classification results (both F1-score and accuracy) are superior for DenseNet while comparing different CNN models used in this study. In addition, the small prediction time along with a fewer number of parameters required to model DenseNet comparing VGG3 shows its efficacy in detecting malware executables from goodware.

5.4 Classification of malware samples into their respective families (i.e., multi-class classification)

We illustrate the results of characterizing malware families on different categories of images. These experiments were repeated on two benchmark datasets used earlier in previous experiments. The accuracy obtained using DenseNet is 98.53%. For Maling executables, an F1-score of 98.88% is attained; refer to Table 4. The confusion matrix of the two datasets is shown in Figs. 10 and 12. From the results, we

can see that 16 out of 25 families attained 100% detection, test samples of 6 malware families had an F1-score greater than 95%, and finally, 92.68% of Malex.gen!j, 89.47% of Swizzor.gen!E, and 42.5% of Swizzor.gen!l were identified by the trained DenseNet model. In particular, we found that 42% of Swizzor.gen!l variants were wrongly labeled as C2LOP.P (refer to Fig. 12).

Figure 11 compares the visual representation of certain samples belonging to three families which reported the highest misclassification (Fig. 12). As we can see, images of malware samples of these families have identical visual patterns due to similar distribution of bytes. Consequently, the classification models wrongly label samples at the prediction time. Thus, to improve the accuracies and F1-score of the DenseNet, we merged malicious samples of C2LOP and Swizzor to form a single-family during the training phase. Specifically, C2LOP is a member of Swizzor; hence, the malicious samples of these families exhibit similar behaviors [49]. The confusion matrix in Fig. 13 shows that the performance of DenseNet significantly improves. Now, the F1-score of malware

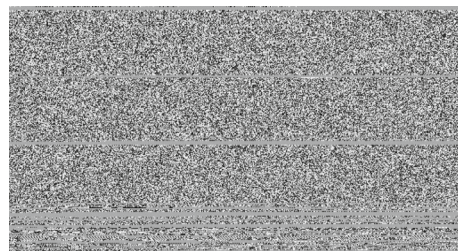
Table 4 Comparison with running time required for Image Dimension 256×256 for BIG 2015 and Maling datasets

Models	A (%)	F (%)	Prediction time
Prior work			
CNN+1LSTM (BIG 2015)	97.64	94.15	32.0 ms
CNN+2LSTM (BIG 2015) [16]	97.91	95.52	62.0 ms
GRU+SVM (Maling) [15]	85.00	85.00	23.7 ms
GIST+KNN (Maling)	92.10	91.70	60.0 ms
GIST+SVM (Maling)	92.50	91.40	64.0 ms
GLCM+SVM (Maling)	93.40	93.00	48.0 ms
Bat algorithm (Maling) [37]	94.60	94.50	20.0 ms
Tuncer et al. [33] (Maling)	88.08	87.62	–
Cayr et al. [34] (Maling)	98.72	96.61	–
Vasan et al. [21] (Maling)	98.82	98.75	0.81 seconds
Kalash et al. [28] (Maling)	98.52	–	–
Kalash et al. [28] (BIG 2015-Setting-A)	98.99	–	–
Khan et al. [31] (BIG 2015)	88.36	–	9248 seconds
Gibert et al. [30] (Maling-25 families)	–	94.8	–
Gibert et al. [30] (Maling-19 families)	–	98.4	–
Gibert et al. [30] (BIG 2015)	97.5	94.0	0.001 seconds
Naeem et al. [32] (Maling)	–	98.75	30 seconds
Our Proposed method: Classification			
Gabor filter applied DenseNet Markov (BIG 2015)	98.52	98.53	11.5 ms
Gabor filter applied DenseNet Markov (Maling)	98.97	98.88	5.0 ms
Gabor filter applied DenseNet Markov images (on combining C2LOP and Swizzor families of Maling)	99.36	99.37	4.95 ms
Our Proposed method: Detection			
Gabor filter applied DenseNet Markov (BIG 2015)	98.98	98.98	5.7 ms
Gabor filter applied DenseNet Markov (Maling)	99.94	99.94	3.8 ms

Fig. 10 Confusion matrix of BIG 2015



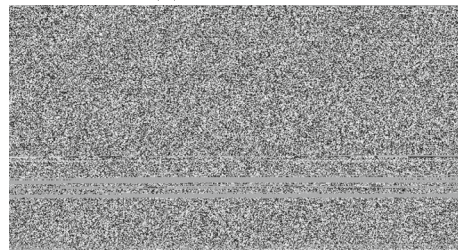
Fig. 11 Visual similarity in samples of C2LOP, Swizzor.gen!E, and Obfuscator.AD families



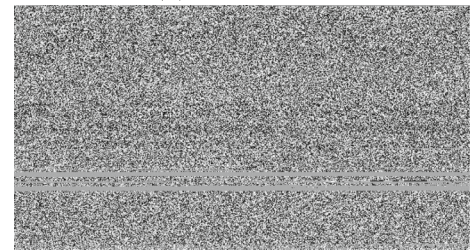
(a) C2LOP.P I1



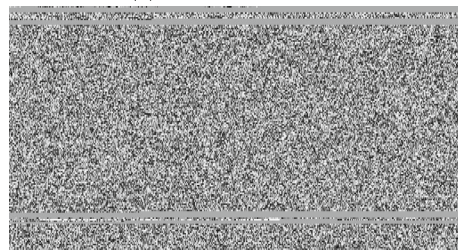
(b) C2LOP.P I2



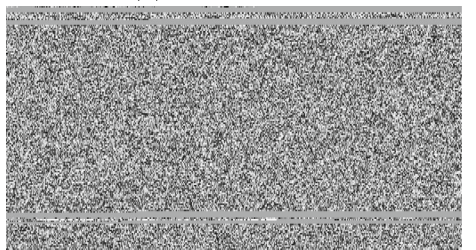
(c) Swizzor.gen!E I1



(d) Swizzor.gen!E I2



(e) Obfuscator.AD I1



(f) Obfuscator.AD I2

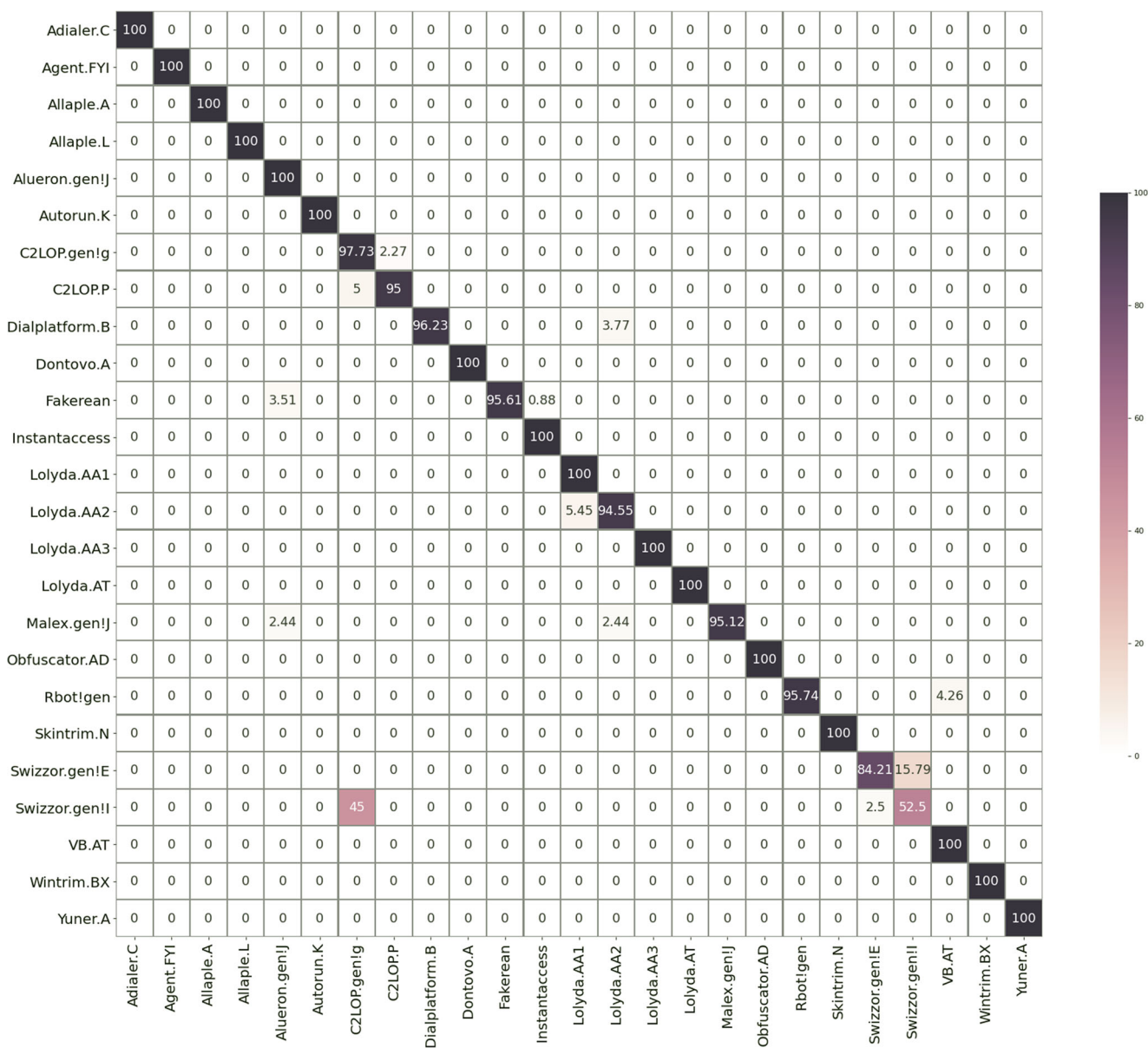


Fig. 12 Confusion matrix of Maling dataset among variants

families is in the range of 95–100%. In particular, we can find that 100% of samples of C2LOP and Swizzor.gen!E were identified by the classification model.

5.5 Comparative analysis of proposed work with the state-of-art solutions

Comparative analysis is performed based on two different experimental settings: malware detection and classification. In [15], an intelligent malware system is proposed on the Maling dataset which achieved an accuracy of 85% and a prediction time of 23.7 ms. Quan et al. [16] proposed a deep learning-based malware classification method that requires only the knowledge of identification of features and

has obtained the maximum accuracy of 97.91%. In [30], BIG2015 and maling datasets were used where the visual similarity was considered and the accuracy obtained was 98.4%. Zhihua et al. [37] give a comparison on various gist techniques and the bat algorithm on the maling dataset and obtained an accuracy of 94.5% with a prediction time of 20 ms.

We compare the outcome of our proposed method with the state-of-the-art solutions. To have fair comparison, we considered only those classification systems which employed identical dataset as we used in our work. Table 4 summarizes the results of prior work and our proposed solution. We can observe that performance of our solutions is better compared to all other works listed in Table 4

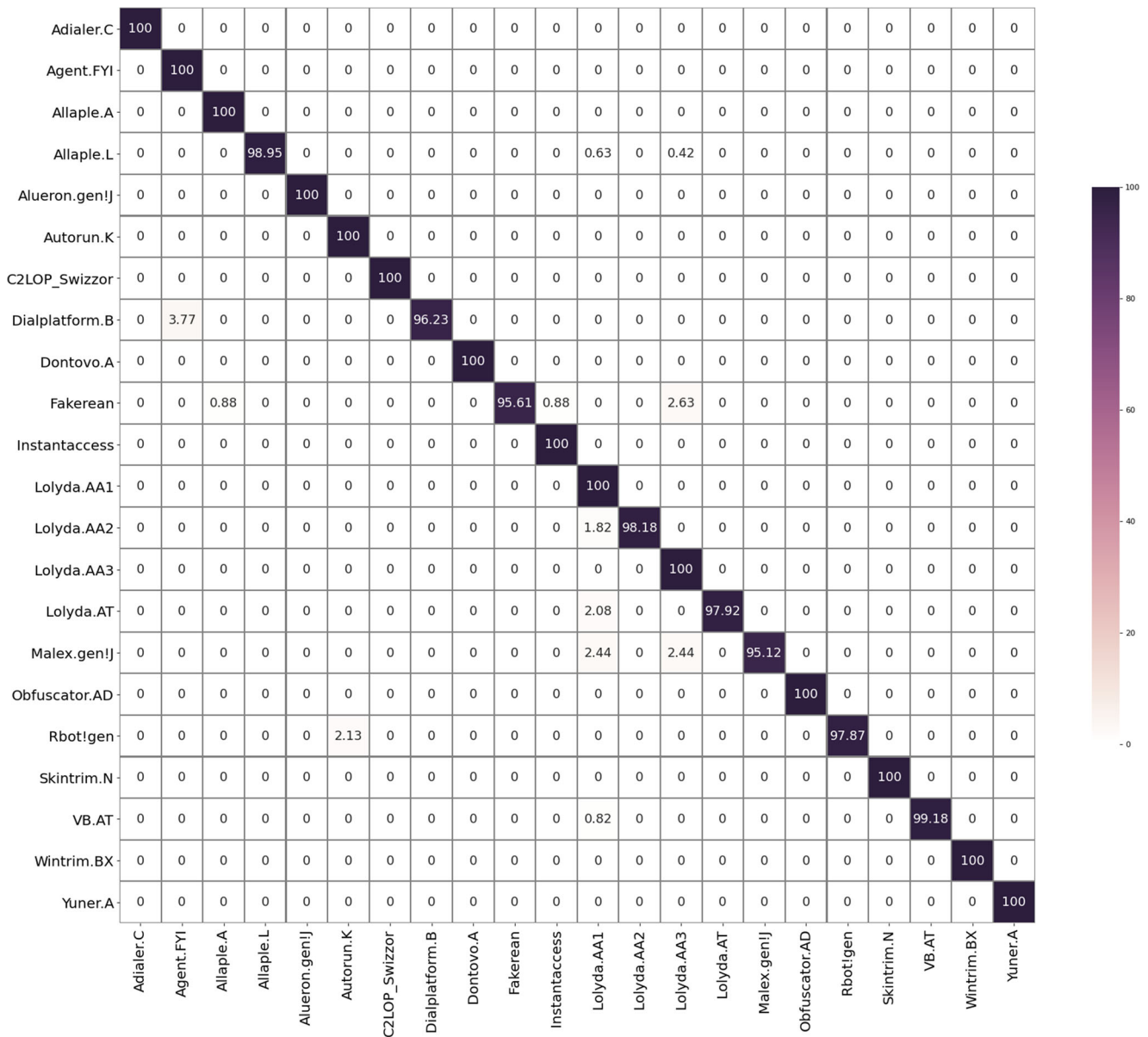


Fig. 13 Confusion matrix combining C2LOP and Swizzor families of Maling dataset

both in terms of values of evaluation metrics and execution time.

5.6 Performance evaluation of DenseNet model on evasion attack: adversarial attacks

To evaluate the robustness of classification models, we performed adversarial attacks. In general, attacks can be categorized as poisoning attacks and evasion attacks. In the former, small perturbations are added to the training samples which forces the classifier to learn a complex hypothesis function. Consequently, the model creates overlapping decision surfaces; eventually, the model wrongly classifies the submitted samples. While in the

evasion attack, samples in the test set are modified to increase the misclassification rate. In this work, we launched an evasion attack with black-box access to classification models. In particular, the adversary has zero knowledge about the classification algorithms and parameters, but may have access to surrogate samples available in public malware repositories.

In our proposed method, all executables are transformed into images; the model is developed by training the classifier on the image textures. For fabricating attacks, we add two types of additive noise (at a noise rate of 0.01) to the samples in the test set. We created evasive samples by injecting (a) Gaussian and (b) Poisson’s noise. A drop in the F1 score obtained is shown in Table 5. The misclassification

Table 5 F1 score of Densenet for Malimg and Big2015 dataset

Noise	Malimg	BIG2015
Poisson's	0.971	0.907
Gaussian	0.97	0.901

is insignificant for Malimg which is a highly imbalanced dataset, but an 8% decline in F1 is obtained for BIG2015 samples. One of the important aspects that were noticed in this dataset is the presence of structural dissimilarity among variants in families, which further increased the addition of noise, causing the classifier to mislabel such instances. In the future, we would like to evaluate the performance of DenseNet by augmenting the training set with adversarial examples using the executable images added with additive noise.

6 Conclusion and future work

In this paper, we developed solutions to detect and classify malicious executables by utilizing the visualization approach. We used a deep convolutional neural network (VGG-3) and DenseNet to extract features from images. Our study analyzes the F1 measure of each model and the best F1 measure is obtained for 256×256 Gabor Markov images. Comprehensive experiments on two public benchmark dataset (Malimg and BIG2015) using DenseNet with Gabor Markov images resulted in 99.94% and 98.88% F1-measure in case of malware detection problem, while F1-measures of 99.37% and 98.88% were obtained for malware classification (family classification). Besides, comparing with the state-of-the-art, we observed improved performance in terms of both detection, classification, and execution time.

In the future, our system can be extended to evaluate the strength of the classification model against the adversarial attacks. This investigation will be conducted by injecting bytes from benign executables in the malware files, which would change the statistical distribution of the malware samples. In order to develop countermeasures on attacks, we will propose a model in which the classification algorithm will be trained with $N + 1$ classes, where N is the original malicious class and one adversarial class. In addition, we would like to investigate generating a classification model using one-shot training techniques, this would identify evolving malware families having fewer malicious variants.

References

1. BBC N Cyber-attack: Europol says it was unprecedented in scale. <http://www.bbc.com/news/world-europe-39907965>
2. <https://blog.logsign.com/the-biggest-cyber-attacks-in-2019/> (accessed Jan 1, 2021)
3. Internet security threat report 2021, <https://www.sophos.com/en-us/medialibrary/pdfs/technical-papers/sophos-2021-threat-report.pdf>
4. <https://crowdresearchpartners.com/portfolio/byod-mobile-security-report/>(accessed Jan.1, 2021)
5. AV-Test Threat Report: <https://www.av-test.org/de/statistiken/malware/> (accessed Jan 1, 2020)
6. Kang H, Jang J, Mohaisen A, Kim HK (2015) Detecting and Classifying Android Malware Using Static Analysis along with Creator Information. *Int J Distrib Sens Netw* **11**(6)
7. Han W, Xue J, Wang Y, Huang L, Kong Z, Mao L (2019) MalDAE: Detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. *Comput Secur* **83**:208–233
8. Pascanu R, Stokes JW, Sanossian H, Marinescu M, Thomas A (2015) Malware classification with recurrent networks. In: *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp 1916–1920.
9. Vinayakumar R, Alazab M, Soman KP, Poornachandran P, Venkatraman S (2019) Robust Intelligent Malware Detection Using Deep Learning. *IEEE Access* **7**:46717–46738
10. Nataraj L, Karthikeyan S, Jacob G, Manjunath BS (2011) Malware Images: Visualization and Automatic Classification. In: *Proc. 8th International Symposium, Visualization for Cyber Security, VizSec*
11. Fu J, Xue J, Wang Y, Liu Z, Shan C (2018) Malware Visualization for Fine-Grained Classification. *IEEE Access* **6**:14510–14523
12. Zhong W, Gu F (2019) A Multi-Level deep learning system for malware detection. *Expert Syst Appl* **133**:151–162
13. Hazra A, Choudhary P, Sheetal Singh M (2021) Recent Advances in Deep Learning Techniques and Its Applications: An Overview. In: Rizvanov A. A, Singh B. K, Ganasala P (eds) *Advances in Biomedical Engineering and Technology, Lecture Notes in Bioengineering*. Springer, Singapore, pp 103–122
14. Gibert D, Mateu C, Planes J, Vicens R (2018) Classification of malware by using structural entropy on convolutional neural networks the thirtieth. *AAAI conference on innovative applications of artificial intelligence (IAAI-18)*
15. Agarap AF, Pepito FJH (2017) Towards building an intelligent anti-malware system: A deep learning approach using support vector machine (SVM) for malware classification, [online] Available: [1801.00318](https://arxiv.org/abs/1801.00318)
16. Le Q, Boydell O, Namee BM, Scanlon M (2018) Deep learning at the shallow end: Malware classification for non-domain experts. *Digit Invest* **26**(Supplement):S118–S126
17. Jain A, Phanishayee A, Mars J, Tang L, Pekhimenko G (2018) Gist: Efficient Data Encoding for Deep Neural Network Training. *ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*, Los Angeles, pp 776–789
18. Zhang J, Qin Z, Yin H, Ou L, Zhang K (2019) A feature-hybrid malware variants detection using CNN based opcode embedding and BPNN based API embedding. *Comput Secur* **84**:376–392
19. Kim T, Kang B, Rho M, Sezer S, Im EG (2019) A multimodal deep learning method for android malware detection using various features. *IEEE Trans Inf Forensic Secur* **14**(3):773–788
20. Nezhadkamali M, Soltani S, Seno SAH (2017) Android malware detection based on overlapping of static features. *7th International Conference on Computer and Knowledge Engineering (ICCKE)*
21. Vasan D, Alazab M, Wassan S, Naeem H, Safaei B, Zheng Q (2020) IMCFN: Image-Based Malware Classification using Fine-tuned Convolutional Neural Network Architecture. *Comput Netw* **171**

22. Shezan FH, Afroze SF, Iqbal A (2017) Vulnerability detection in recent Android apps: An empirical study. 2017 International Conference on Networking, Systems and Security (NSysS)
23. Xue D, Li J, Lv T, Wu W, Wang J (2019) Malware Classification Using Probability Scoring and Machine Learning. *IEEE Access* 7:91641–91656
24. Yoo S, Kim S, Kim S, Kang BB (2020) AI-hydra: Advanced hybrid approach using random forest and deep learning for malware classification. *Inf Sci* 546:420–435
25. Zhao Y-I, Qian Q (2018) Android Malware Identification Through Visual Exploration of Disassembly Files. *Int J Netw Secur* 20(6):1061–1073
26. Han K, Kang B, Im EG (2014) Malware analysis using visualized image matrices. *Sci World J*:1–15
27. Yuan B, Wang J, Liu D, Guo W, Wua P, Bao X (2020) Byte-level Malware Classification Based on Markov Images and Deep Learning. *Comput Secur* 92
28. Kalash M, Roohan M, Mohammed N, Bruce NDB, Wang Y, Iqbal F (2018) Malware Classification with Deep Convolutional Neural Networks. 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, pp 1–5
29. Roseline SA, Geetha S, Kadry S, Nam Y (2020) Intelligent Vision-Based Malware Detection and Classification Using Deep Random Forest Paradigm. *IEEE Access* 8:206303–206324
30. Gibert D, Mateu C, Planes J, Vicens R (2019) Using convolutional neural networks for classification of malware represented as images. *J Comput Virol Hacking Techn* 15(1):15–28
31. Khan RU, Zhang X, Kumar R (2019) Analysis of ResNet and GoogleNet models for malware detection. *J Comput Virol Hacking Techn* 15(1):29–37
32. Naem H, Ullah F, Naem MR, Khalid S, Vasan D, Jabbar S, Saeed S (2020) Malware detection in industrial internet of things based on hybrid image visualization and deep learning model. *Ad Hoc Netw* 10:102154
33. Tuncer T, Ertam F, Dogan S (2021) Automated malware identification method using image descriptors and singular value decomposition. *Multimed Tools Appl*:1–20
34. Çayır A, Unal U, Dağ H (2021) Random CapsNet forest model for imbalanced malware type classification task. *Comput Secur* 102:102133
35. Sun G, Qian Q (2021) Deep Learning and Visualization for Identifying Malware Families. *IEEE Trans Depend Sec Comput* 18(1):283–295
36. Feng R, Chen S, Xie X, Meng G, Lin S-W, Liu Y (2021) A Performance-Sensitive Malware Detection System Using Deep Learning on Mobile Devices. *IEEE Trans Inf Forensic Secur* 16:1563–1578
37. Cui Z, Xue F, Cai X, Cao Y, Wang G-g, Chen J (2018) Detection of Malicious Code Variants Based on Deep Learning. *IEEE Trans Ind Inform* 14(7)
38. Portable freeware collection, <https://www.portablefreeware.com/> (accessed Feb 1, 2020)
39. Softonic, <https://en.softonic.com/windows> (accessed March 1, 2020)
40. Sourceforge <https://sourceforge.net/> (accessed March 1, 2020)
41. DriverPack solution, <https://drp.su/en> (accessed February 18, 2020)
42. Ronen R, Radu M, Feuerstein C, Yom-Tov E, Ahmadi M (2018) Microsoft Malware Classification Challenge, arXiv: [1802.10135](https://arxiv.org/abs/1802.10135)
43. Yajamanam S, Selvin VRS (2018) Fabio di troia and Mark Stamp. Deep Learning versus Gist Descriptors for Image-based Malware Classification. 4th International Conference on Information Systems Security and Privacy, pp 553–561
44. Huang G, Liu Z, van der ML, Weinberger KQ (2017) Densely connected convolutional networks 2017. IEEE conference on computer vision and pattern recognition (CVPR)
45. Hussain M, Jordan J, Bird JJ, Faria DR (2018) A Study on CNN Transfer Learning for Image Classification. Proceedings of 18th Annual UK Workshop on Computational Intelligence Nottingham
46. Rezende E, Ruppert G, Carvalho T, Theophilo A, Ramos F, de Geus P (2018) Malicious software classification using VGG16 deep neural network's bottleneck features. In: Information Technology-New Generations. Springer, Cham, pp 51–59
47. Hazra A (2021) A comprehensive survey on chest diseases analysis: technique, challenges and future research directions. *International Journal of Multimedia Information Retrieval*
48. Choudhary P, Hazra A (2019) Chest disease radiography in twofold: using convolutional neural networks and transfer learning. *Evolving Systems*
49. C2LOP: <https://www.microsoft.com/en-us/wdsi/threats/malware-encyclopedia-description?Name=Trojan:Win32/C2Lop.gen!M> (Accessed on 2-December,2020)

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

Optimal Distribution of Workloads in Cloud-Fog Architecture in Intelligent Vehicular Networks

Mahdi Abbasi¹, Mina Yaghoobikia, Milad Rafiee², Mohammad R. Khosravi³,
and Varun G. Menon⁴, *Senior Member, IEEE*

Abstract—With the fast growth in network-connected vehicular devices, the Internet of Vehicles (IoV) has many advances in terms of size and speed for Intelligent Transportation System (ITS) applications. As a result, the amount of produced data and computational loads has increased intensely. A solution to handle the vast volume of workload has been traditionally cloud computing such that a substantial delay is encountered in the processing of workload, and this has made a serious challenge in the ITS management and workload distribution. Processing a part of workloads at the edge-systems of the vehicular network can reduce the processing delay while striking energy restrictions by migrating the mission of handling workloads from powerful servers of the cloud to the edge systems with limited computing resources at the same time. Therefore, a fair distribution method is required that can evenly distribute the workloads between the powerful data centers and the light computing systems at the edge of the vehicular network. In this paper, a kind of Genetic Algorithm (GA) is exploited to optimize the power consumption of edge systems and reduce delays in the processing of workloads simultaneously. By considering the battery depreciation, the supporting power supply, and the delay, the proposed method can distribute the workloads more evenly between cloud and fog servers so that the processing delay decreases significantly. Also, in comparison with the existing methods, the proposed algorithm performs significantly better in both using green energy for recharging the fog server batteries and reducing the delay in processing data.

Index Terms—Cloud, fog, genetic algorithm, Internet of vehicles, workload allocation.

I. INTRODUCTION

AS a result of the tremendous growth in the number of smart vehicular devices, the Internet of Vehicles (IoV) has experienced rapid expansion. The increase in the number of devices has caused a multiplication of data and large-scale computation loads [1], [2]. Cloud computing has been proposed as a solution to manage these loads [3]. However,

Manuscript received May 7, 2020; revised October 13, 2020 and February 13, 2021; accepted April 2, 2021. Date of publication April 19, 2021; date of current version July 12, 2021. The Associate Editor for this article was A. Jolfaei. (*Corresponding author: Mahdi Abbasi.*)

Mahdi Abbasi, Mina Yaghoobikia, and Milad Rafiee are with the Department of Computer Engineering, Faculty of Engineering, Bu-Ali Sina University, Hamedan 65178-38695, Iran (e-mail: abbasi@basu.ac.ir; m.yaghoobikia@eng.basu.ac.ir; m.rafee@alumni.basu.ac.ir).

Mohammad R. Khosravi is with the Department of Computer Engineering, Persian Gulf University, Bushehr 75169-13817, Iran, and also with the Department of Electrical and Electronics Engineering, Shiraz University of Technology, Shiraz 71557-13876, Iran (e-mail: m.khosravi@sutech.ac.ir).

Varun G. Menon is with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Kochi 683582, India (e-mail: varunmenon@scmsgroup.org).

Digital Object Identifier 10.1109/TITS.2021.3071328

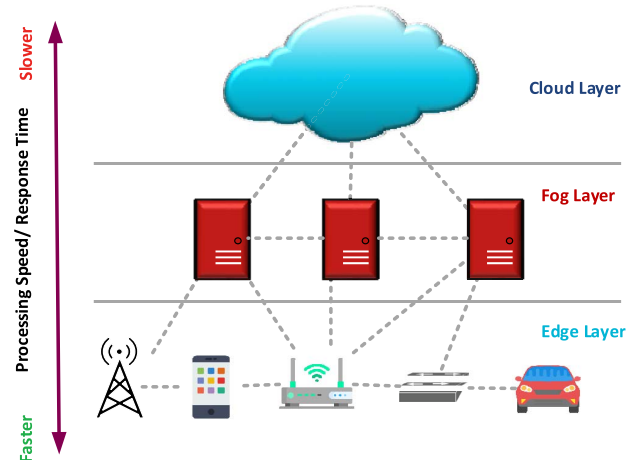


Fig. 1. IoV data processing layer stack.

the time-consuming nature of the processing of workloads in clouds is still a major issue in the field of distributed vehicular networks [4]. Processing the workloads at the edges of the vehicular network can reduce the processing time, but the transmission of workloads from the data centers (which are equipped with sustainable electric power) to the edges leads to serious limitations in terms of supplying the required power for computing [5], [6]. Thus, we need to achieve a balance in distributing the processing requests between the cloud and the edge [7].

Fig. 1 shows the layers of data processing in a cloud-fog architecture for IoV. As can be seen in the figure, the lowest layer contains vehicular devices that produce the data. These devices can use their own processing resources and process the data in positions close to the user. Although the proximity of edge devices to the end-user remarkably reduces the delay in request transmission and increases the response rate, these devices have a lower processing power than the cloud. In the next layer lie powerful routers and servers which are close to the edge and can process the workloads without transferring them to the cloud [8]. By moving away from the edge of the Internet to towards the data centers, the transmission delay will increase. In the highest layer, large data centers that provide the enhanced capability of processing and storage are distributed as clouds around the world. As they are very far from the end-users, these resources usually impose long delays in the processing of requests [9]. Also, they also consume high amounts of electric power, whereas most edge devices can function with small amounts of power or even with batteries.

Fog computing is a kind of distributed computing that can replace cloud computing by using several devices near the edge of the vehicular network (see Fig. 1). Fog computing is more efficient than edge computing in terms of processing, while it is less potent than cloud computing. The chief issue in fog computing is the high costs of the required electric power. Nowadays, a more challenging problem is providing sustainable energy resources that can afford the long-term energy requirements of fog nodes in IoV [10], [11]. The processing nodes chiefly receive their required power from rechargeable batteries [12]. As this type of power source is extremely limited and should be frequently recharged, the use of renewable energies as a secondary or even the only power supply at the network edge is necessary [13]–[15]. Thus, we need to develop a method for striking a balance in distributing the workloads among fog nodes and cloud data centers so that both delay and power consumption could be optimized. As a result, the energy resources of the IoV become more sustainable.

To achieve this goal, the present study makes use of a genetic algorithm in finding the best distribution for the workloads. A review of the literature indicates that few studies have addressed the issue of finding the best cost function and the effect of the coefficients of this function on the algorithm's decision-making. Given this, this study first introduces a cost function of distribution based on two parameters, i.e., power and delay, and then attempts to modify the coefficients corresponding to these parameters according to a genetic algorithm in order to attain the best coefficients of workload distribution in a way that the workloads could be processed with the least delay and the least amount of power consumption.

The genetic algorithm is a method for finding approximate solutions to search and optimization problems. This algorithm is considered as a kind of evolutionary algorithm due to its use of biological concepts such as inheritance and mutation. Genetics addresses inheritance and the transfer of attributes from one generation to the next. In living creatures, chromosomes and genes are responsible for this transfer. This mechanism acts in a way that superior and stronger chromosomes will survive. The final result is that stronger creatures would be able to survive. Genetic programming is a technique of programming that uses genetic evolution as a model for problem-solving. Over time, the genetic algorithm has grown in popularity in a diversity of problems such as optimization, image processing, topology, artificial neural network training, and decision-making systems [16].

A genetic algorithm begins with initializing a random population, which is composed of the possible solutions to the problem. Each solution is a chromosome, and the entire chromosomes form the initial population. In the first step, the value of each chromosome in the population is specified by the fitness function. During the execution of the algorithm, parents with more fitness are selected for reproduction, and the next generation is generated using genetic operators. Crossover, mutation, and selection are the three main operators in genetic algorithms [17].

By using a genetic algorithm, the present paper seeks to obtain the best value function so that we could strike a balance between power consumption at the edge of network and delay

in the transmission of workloads as well as minimize these two parameters. Also, we use renewable energies in the processing and transmission of workloads at the network edge due to the significance of sustainable energy resources in computing tasks. Another innovation of this study is its use of renewable energy as an input parameter in the genetic algorithm. For this purpose, green energy is used in our proposed method to calculate the value function of the algorithm. The main reason for using renewable energies is the limitation of edge devices on power consumption. As a result, these devices need to be regularly connected to a power source for being recharged, which limits their mobility. Also, changing the battery in IoV devices may impose high costs and is sometimes dangerous. For this reason, IoV devices should be able to maintain their independence and sustainability by using green energies and wireless charging ability [18]. In this vein, we aim to utilize renewable energies to minimize the number of batteries at the network edge.

The structure of the paper is as follows. Section 2 is a review of the related literature. In Section 3, the workload allocation model is formulated. Next, after a brief description of the structure of the genetic algorithm, the optimization of delay and power consumption in a cloud-fog environment is discussed. Section 5 describes the implementation of the proposed method and evaluates it in terms of the parameters of the algorithm. The method is also compared with other existing methods. Finally, some concluding remarks are made and ideas for further research are suggested.

II. REVIEW OF LITERATURE

In recent years, many researchers have studied the methodologies to orchestrate the distribution of workloads and reduce the overall processing delay in IoV. We briefly review some of the recent studies that have investigated the optimization of energy usage and delay reduction in fog computing.

Pioneering work has been presented by Xu and Ren [19]. In this work, they inspect the possibility of using renewable energies as backup energy sources in mobile edge networks. Their method uses machine learning algorithms to manage the energy resources and distribute the computation workloads. Their method aims to minimize the prominent costs of the processing requests that include the processing delay and consuming energy. The consequence of using a slow learning mechanism in their method is the weak results in controlling the power consumption of edge nodes.

Unfortunately, in many of the recently proposed methods, only one aspect of the problem is considered. That is, some of them sacrifice the processing time for optimizing the power consumption at edge systems, and vice versa. Hence, in any method to be developed, the processing delay and consuming power should be optimized simultaneously.

Regarding abovementioned challenge, Xu *et al.* presented a reinforcement learning-based method [20]. Their algorithm was able to learn and adapt itself to any system with unknown modeling parameters. Despite the undeniable results of their method in achieving acceptable performance in orchestrating the edge computations and using renewable energy sources for mobile edge nodes, their algorithm failed to fairly distribute the workloads among the computing nodes.

The GLOBE method of Wu *et al.* [12] tries to optimize the performance processing nodes at the network edge by geographically balancing the distribution of loads, and at the same time, controlling the input load of any edge nodes. This method can handle stochastic events concerning the battery status and power limitations. Although the GLOBE is slightly successful in optimizing the battery energy level, it is still so far from perfect.

In 2019, Dalvand and Zamanifar [21] proposed a new model for processing data in the Internet of things (IoT) and developed an IoT-Fog-Cloud in which the fog layer is geographically close to the IoT edge devices. In their system, a multi-purpose dynamic service is created to achieve a balance between delay and resource costs. This service is formulated by MILP and solved through weighted goal programming. The method controls and minimizes only one goal as a compromise between time and power consumption.

None of the above studies have been able to strike a balance between cost and power consumption in distributing workloads among network nodes. Our work is aimed at developing a mechanism for the balanced distribution of workloads between the cloud and edge nodes. This mechanism is supposed to attain a desirable tradeoff between power consumption and workload delay. It will make use of renewable energy sources as the power supply for edge computations. These sources are expected to preserve the battery charge level.

III. THE PROPOSED METHOD

Our point of departure is the fact that none of the previous works have offered an optimum solution to reduce the costs of fog computing. To elaborate on our proposed method and evaluate it, a cloud-fog environment is simulated as in [10]. This environment involves four main models: workload, delay, power consumption, and battery status. In the following, we shall first examine these models as described in the reference and then present our proposed algorithm. Next, we will define a new scenario to study how the algorithm functions. The proposed scenario will be simulated for evaluation.

A. Formulation of the Problem

For the chief scenario, the edge system includes a base station and a set of edge servers that are set geographically close to each other. A battery with a limited capacity is used in each computing resource at the network edge (i.e., fog servers). The shared power supply mechanism used in the network lets the workloads be sent to the cloud especially when the edge servers require battery charge. The workload sent by the users to the edge is first received by the base station. The base station manages and decides on the amount of workload that must be allocated to the edge or transmitted to the cloud. The definition of the formulation parameters is presented in Table I.

The proposed system is modeled here by considering it from four aspects.

1) *Workload*: The equal time intervals $t = 0, 1, 2, \dots$ are used to model the time. The computation capacity of each edge device is specified in each time interval in terms of the number of active servers; $\lambda(t) \in [0, \lambda_{max}]$ is the rate of assigning the workloads to edge nodes and $\mu(t)$

TABLE I
THE MAIN SYMBOLS

Symbol	Meaning	Symbol	Meaning
$\lambda(t)$	The total rate of the input load	$c_{delay}(t)$	The total cost of delays
$\mu(t)$	The amount of workload processed locally	$c_{back}(t)$	The cost of using the supporting power supply
$m(t)$	The number of active servers at the network edge	$d_{op}(t)$	Power consumption for operational tasks at the network edge
$c_{io}(t)$	The cost of delay in processing workloads at the network edge	$c_{comp}(t)$	Power consumption for processing loads at the network edge
$c_{off}(t)$	The cost of delay in sending workloads to the cloud	$d(t)$	Total power consumption
$g(t)$	The total renewable power received	$b(t)$	Battery level at the network edge
$h(t)$	The congestion status of the network	$s(t)$	System status

is the rate by which the edge nodes process the assigned workload. Finally, $\lambda(t) - \mu(t)$ denotes the remaining part of the workload transmitted to the cloud. The number of dynamic servers in any time interval is $m(t) \in [0, M]$. This number may change in different time intervals.

2) *Delay*: We consider three different delays in the system model:

2-1. The delay in communicating workloads on the wireless network, which is shown by $c_{wi}(t)$. This delay depends on the input load of the network (i.e., $\lambda(t)$). In our model, it is assumed as 0 due to the physical closeness of the active nodes.

2-2. The delay of processing workloads at local subregions of the network edge, which is shown by $c_{lo}(t)$. The amount of this delay directly depends on the number of active servers, the processing rate of them, and the model of managing queues in each of them. In our experiments, the M/G/1 mechanism models the queue management in any active server running on edge nodes. As a result, the delay in processing at the network edge is estimated using the following equation [22]:

$$c_{lo}(\mu(t), m(t)) = \frac{\mu(t)}{m(t) \cdot k - \mu(t)} \quad (1)$$

In this equation, kk represents the processing capacity of each active server.

2-3. The delay in communicating the residual workload to the cloud is shown by $c_{off}(t)$ and is estimated based on the congestion status of the network. This status is represented by $h(t)$, and is computed by adding the round-trip time (RTT) delay and the processing delay of the cloud. As a result, this delay is calculated based on $h(t)$ according to the following equation [22]:

$$c_{off}(h(t), \mu(t), \lambda(t)) = (\lambda(t) - \mu(t)) h(t) \quad (2)$$

Finally, the cost of overall delay of the aggregate input workload is estimated by adding the three above delays [22]:

$$\begin{aligned} c_{delay}(h(t), \lambda(t), \mu(t), m(t)) \\ = c_{wi}(\lambda(t)) \\ + c_{lo}(\mu(t), m(t)) + c_{off}(h(t), \mu(t), \lambda(t)) \end{aligned} \quad (3)$$

Note that, $c_{wi}(\lambda(t))$ is negligible.

3) *Power Consumption*: The total consumed power is composed of two parts:

3-1. A part of the power is used for basic operations and communicating the loads. This part is represented by $d_{op}(t)$ and is independent of any operations regarding

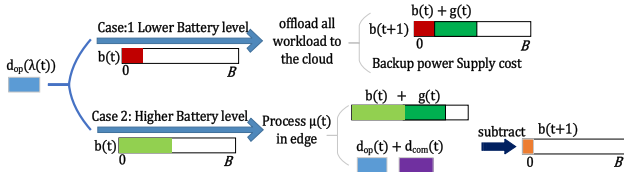


Fig. 2. Two modes of battery status.

processing the loads but merely depends on the input load of the network ($\lambda(t)$). In our model, $d_{op}(t)$ is composed of two power types [22]:

$$d_{op}(\lambda(t)) = d_{sta} + d_{dyn}(\lambda(t)) \quad (4)$$

The d_{sta} and $d_{dyn}(\lambda(t))$ represent the static power consumption of the network edge and the dynamic power consumption, respectively. The latter differs from the input load of the network and is set to 0 in our model due to the physical closeness of the computing nodes.

3-2. The power required for the processing of workloads at the edge is shown by $d_{com}(t)$. To estimate this parameter, the amount of the workload allocated to the edge ($\mu(t)$) and the number of active edge servers ($m(t)$) are required. Finally, the total required power is obtained by the following equation [22]:

$$d(\lambda(t), \mu(t), m(t)) = d_{op}(\lambda(t)) + d_{com}(\mu(t), m(t)) \quad (5)$$

In this model, $g(t)$ denotes any renewable energy source that can be used as the power supply $g(t)$.

4) Battery Status: As formerly explained, one battery with limited charge is used to supply each active edge server. Overall, the total battery charge at the network edge is $b(t) \in [0, B]$, where B denotes a predefined maximum capacity. The renewable energy sources can recharge these batteries. The initial battery level is set to 0. To control the battery level at the network edge, we should control the rate of processing of workloads at the edge servers. Hence, the state of the battery is determined by the following conditions:

D-1. When $b(t) \leq d_{op}(\lambda(t))$, no processing is allowed at the network edge. In this state, since the battery charge is not sufficient, the whole workload $\lambda(t)$ is transmitted to the cloud. In this state, the renewable energy sources recharge the battery. The overall cost of communicating the workload to the cloud is calculated by the following equation [22]:

$$c_{bak}(\lambda(t)) = \varphi \cdot d_{op}(\lambda(t)) \quad (6)$$

where φ is the coefficient reflecting the cost of consuming the supporting power supply. In the next interval, the renewable power source will charge the battery according to equation (7) [22].

$$b(t+1) = b(t) + g(t) \quad (7)$$

The first state in Fig. 2 shows this state.

D-2. If the battery level is sufficiently more than the required power for processing a part of workload, that part of the workload ($\mu(t)$) is processed at the edge, and the remaining part ($\lambda(t) - \mu(t)$) is transmitted to the cloud. Thus,

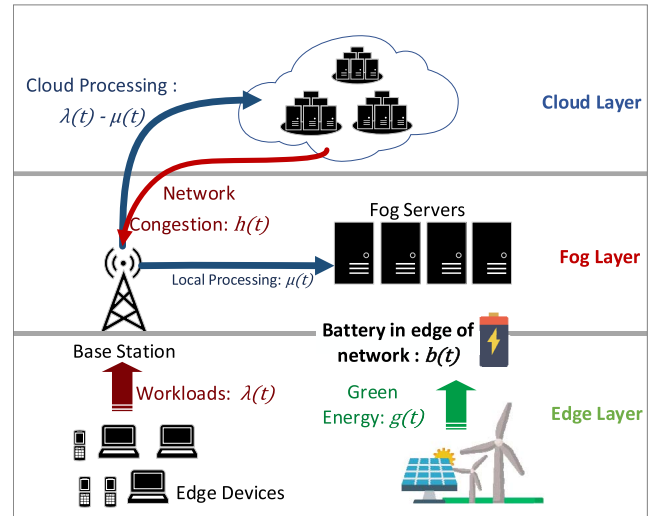


Fig. 3. The cloud-fog architecture.

the following equation calculates the battery level in the next interval as:

$$b(t+1) = b(t) + g(t) - d(\lambda(t), \mu(t), m(t)) \quad (8)$$

The operational cost of the battery in this state is:

$$c_{battery}(t) = \omega \cdot \max\{d(\lambda(t), \mu(t), m(t)) - g(t), 0\} \quad (9)$$

where $\omega > 0$ is the operating cost of one battery unit.

This state is shown by the second state in Fig. 2. Based on the above four models, the architecture of the proposed system can be illustrated as in Fig. 3.

In this figure, the set of requests $\lambda(t)$ which have been sent by the users enter the base station. The base station is responsible for distributing the loads between edge servers (i.e., fog servers) and the cloud. The base station uses the evolutionary algorithm to calculate the amount of workload that can be processed by edge servers ($\mu(t)$). Then, the excessive requests are transmitted to the cloud. The transmission of workloads to the cloud creates congestion in the network and imposes longer delays on the loads. Therefore, the congestion is measured in every interval ($h(t)$) to be taken into account in subsequent decisions. In the meantime, renewable energy sources ($g(t)$) provide the power required for edge computations in each interval. If renewable sources produce more energy than is needed by the servers, the surplus is stored in network batteries $b(t)$. Conversely, if the produced renewable energy is inadequate, the batteries will be used.

IV. USING GENETIC ALGORITHM IN THE OPTIMIZATION OF WORKLOAD DISTRIBUTION

This section describes how a genetic algorithm can be used to distribute the workloads more efficiently. The aim of using a genetic algorithm is to minimize system costs. The solution to this problem using a genetic algorithm is presented in Algorithm 1. Below is the description of the algorithm.

In the beginning, the battery level is checked. If the battery level is not sufficient for the basic operation, the supporting power supply is used, and the entire input workload is transmitted to the cloud. In this case, $\mu(t) = 0$, and the genetic

algorithm is not executed (lines 1 and 2). However, if the battery level is high enough for the basic operation to run, all or part of the input load can be processed at the network edge. In this case, the genetic algorithm is used to calculate $\mu(t)$ (lines 3 to 29). In the first step of the genetic algorithm, the initial population is generated (line 4). This population consists of a set of chromosomes. Each chromosome indicates the amount of workload that can be computed at the edge. Next, the fitness of the initial population is calculated (line 5).

The fitness function returns a non-negative value for each chromosome which is indicative of the individual capacity of that chromosome to reduce the costs. The cost function [20] can be used to calculate the fitness of a chromosome. The proposed algorithm attempts to reduce this amount in order to minimize system costs. Given the battery status of the system, the cost function can be calculated in two ways:

$$c(t) = c_{delay}(h(t), \lambda(t), 0, 0) + c_{bak}(\lambda(t)),$$

$$if(b(t) \leq d_p(\lambda(t)))$$

$$c(t) = c_{delay}(h(t), \lambda(t), \mu(t), m(t)) + c_{battery}(t),$$

$$else \quad (10)$$

This equation is composed of two parts: delay cost and power cost. The following two coefficients are used for the power cost part:

- 1) Battery depreciation coefficient (ω)
- 2) Cost coefficient of the supporting power supply (φ)

As the effect of delay is directly involved in the cost function, a new coefficient called delay cost coefficient (θ) is introduced. The proposed algorithm modifies these coefficients to examine their effect on power consumption and workload delay and to find the optimum state on the network.

Another important genetic operator is crossover. Crossover is used to exchange information between two chromosomes, which accelerates convergence in the genetic algorithm. The probability of the effectiveness of this operator lies in the range of 0.6-0.9. This value is called a crossover rate and denoted by $P_{crossover}$. In this problem, two parents and a random position in the parents' genes are selected. Next, the genes on the right side of the random position of the first parent and those on the left side of the random position of the second parent are selected to produce a new chromosome (lines 9 to 15). Another operator is the mutation, which is responsible for producing new information. This operator randomly changes one of the genes of the child with a low probability, such as 0.01. The probability of mutating any chromosome is called the mutation rate and is denoted by $P_{mutation}$. In the proposed algorithm, one gene from the chromosome is randomly selected and changed (lines 16-19). In this algorithm, the number of children produced by crossover and mutation is set by the variable N_c . In each step of this operation, a new child is added to the set P (line 20). Then the fitness function of the generated population is obtained by crossover and mutation operators as was done for the initial population (line 22).

There are different methods in genetic algorithms to select the superior chromosome and transfer it to the next generation. One of the common methods is tournament selection [23]. In this method, two chromosomes are randomly selected from the population. Next, a random number r between 0 and 1

Algorithm 1 Using a Genetic Algorithm in the Optimization of the Workload Distribution

Input : $\lambda, g, h, b, N_c, N_g, P_{crossover}, P_{mutation}, P_{selection}$
Output : μ

```

1: if  $b(t) \leq d_p(\lambda(t))$ 
2:    $\mu(t) \leftarrow 0$ 
3: else
4:    $P \leftarrow Create\ Population()$ 
5:    $fitness(P)$ 
6:   do
7:     for  $i \leftarrow 0, 1, \dots, N_c // crossover\ and\ mutation$ 
8:        $parent1 \leftarrow random(P)$ 
9:        $parent2 \leftarrow random(P)$ 
10:       $child \leftarrow parent1$ 
11:      if  $(random() > P_{crossover})$ 
12:         $point \leftarrow$ 
13:           $random(length\ of\ chromosome)$ 
14:           $child \leftarrow$ 
15:             $crossover(parent1, parent2, point)$ 
16:      End if
17:      if  $(random() > P_{mutation})$ 
18:         $gen \leftarrow random(length\ of\ chromosome)$ 
19:         $child(gen) \leftarrow mutation()$ 
20:      End if
21:       $Add\ a\ child\ to\ P$ 
22:    end for
23:     $fitness(children\ created\ by\ crossover\ section)$ 
24:     $P \leftarrow selection(P, P_{selection})$ 
25:  while
26:     $\mu(t) \leftarrow best\_chromosome(P)$ 
27:    while
28:       $(battery + green < PowerConsumption(\mu(t)))$ 
29:       $\mu(t) \leftarrow next\_best\_chromosome(P)$ 
30:    end while
31: End if

```

is generated. If $r < P_{selection}$ ($P_{selection}$ is a parameter, e.g., 0.8), the fitter individual will be selected as the parent; otherwise, the less fit individual will be selected. These two are again returned to the population and involved in the selection process. After the selection process, the selected chromosomes are introduced as the new generation and sent to the next iteration of the algorithm (line 23). In the proposed genetic algorithm, the child generation operators such as crossover and mutation as well as fitness calculation and selection are executed for N_g times, which is indicative of the number of generations (line 24). When all generations have been executed, the first element of the population will be put in $\mu(t)$ as the final result (line 25).

If the selected chromosome (which indicates the distribution of processable load at the network edge $\mu(t)$) faces battery limitations, the next chromosome in the population should be selected. The process will continue until the power consumption for $\mu(t)$ becomes proportional to the edge batteries (lines 26-28). At the end of the algorithm, the best value is selected for $\mu(t)$, which in addition to minimizing the cost of delay and

Algorithm 2 The Effect of ω and θ on the Proposed Method

Input : λ, g, h **Output** : average delay, average power consumption

```

1: for  $\theta \leftarrow 0.01$  to 1 step 0.01 do
2:   for  $\omega \leftarrow 0.01$  to 1 step 0.01 do
3:     GA_Algorithm( $\lambda, g, h, \theta, \omega$ )
4:   End for
5: End for

```

power consumption regulates power consumption according to the level of edge batteries.

V. IMPLEMENTATION AND EVALUATION

This section describes the implementation and evaluation of the proposed method for optimum distribution of workloads between the cloud and the fog. For this purpose, the evaluation parameters of the problem, the parameters of the different genetic operators, and the implementation environment are examined. Next, the effect of the variations in ω , and θ on the distribution of workloads is studied and the optimum value of these two parameters is obtained. Finally, the proposed method with the optimum values of ω and θ is compared with other existing methods.

A. Simulation Parameters

This section describes the simulation of a cloud-fog environment in order to evaluate the proposed method. In this environment, the genetic algorithm described above is used in the base station as the distributor of workloads between the cloud and fog servers. The simulation aims to examine the effect of the delay cost coefficient and battery depreciation coefficient on the fitness function as well as on the average delay in workload transmission and the power consumption at the network edge. To narrow down the search space in the genetic algorithm, we assume the cost coefficient of the supporting power supply (0.15) as constant and only study the variations in ω , and θ . The process is shown in Algorithm 2. According to this algorithm, with changing the value of ω and θ , the genetic algorithm runs 10000 times in each experiment and the average energy consumption and the delay are measured. In these experiments, $0.01 \leq \omega \leq 1$ and $0.01 \leq \theta \leq 1$, and their values are changed by 0.01 in each experiment.

The proposed method was examined on a system with an 8-core 1.8 GHz CPU and 12GB RAM. In the following, we first initialize the parameters and then discuss the results. The amount of input workload in each interval is specified by a random number that uniformly varies between 10 and 100 requests per second. The renewable energy fluctuates according to a normal distribution of $N(520W, 150)$ [20]. The maximum capacity of each battery is $B = 2kWh$. Also, we assume that the initial charge of battery $b(0) = 0$. The static power consumption of the base station is $d_{sta} = 300W$. We set the maximum number of edge servers $M = 10$. Also, each active server consumes 150W of electricity. The maximum processing rate of each server is 20 requests

per second. We restrict the maximum number of generations of our evolutionary algorithm to 100.

B. The Effect of ω and θ on Workload Distribution

In this section, the results of the experiments are presented using graphs. Then the graphs are analyzed and, by normalizing the values of delay and power consumption, the best coefficients of the fitness function to minimize the costs are obtained.

Fig. 4 illustrates the average delay cost in different experiments for ω and θ . As can be seen in Fig. 4(a), the increased delay coefficient decreases the average delay cost. The reason behind this decrease is the stronger effect of θ on the cost function, which the genetic algorithm seeks to reduce. In fact, the system attempts to reduce the delay cost so that more workloads could be processed locally. For example, Fig. 4(b) shows the variations in the average delay depending on the varying values of θ . In this figure, assuming a constant coefficient of battery depreciation ($\omega = 1$), an increase in the delay coefficient results in a decrease in the delay cost. The most important reason behind the decrease in the delay is the increased value of this parameter in the fitness function as well as the processing of increased amounts of workloads in the fog servers.

Fig. 4(c) depicts the average delay according to the variations of ω for two constant values of θ . When $\theta = 0.01$ (the minimum value), the majority of processes are conducted in the cloud, and the average delay is maximized due to the minimal effect of this parameter on the fitness function and the decision-making. It can be seen that when the delay coefficient is constant, by increasing the battery depreciation coefficient (ω) from 0.01 to 1, the power consumption part in the cost function becomes more significant. Therefore, the system attempts to send more workloads to the cloud to reduce power consumption. Consequently, with the transmission of the loads to the cloud, the average delay begins to escalate. Also, a comparison of the two lines depicted in the figure would show that the average delay with $\theta = 0.08$ is less than with $\theta = 0.01$, which can be explained by its increased effect on the fitness function. Given the above discussion, the greater the coefficient of battery depreciation (ω) and the greater the delay coefficient (θ), the less the average delay.

Fig. 5(a) shows the average power consumption with ω and θ in each experiment. In general, an increase in the coefficient of battery depreciation will reduce power consumption. The reason for this reduction is the increased effect of battery depreciation on the cost function. In fact, the algorithm tries to allocate most of the processes to the cloud to reduce power consumption in the fog servers. Fig. 5(b) depicts the average power consumption with three constant values of θ according to the variations of ω . It can be observed that, as the coefficient of battery depreciation increases, the average power consumption with $\theta = 0.15$ and $\theta = 0.01$ is reduced from 550 w to 450 w. The reason for this reduction is the system's attempt to send more workloads to the cloud and decrease power consumption in the fog servers. As the figure shows, in points where $\theta = 0.01$ (i.e., the minimum value), the majority of workloads are sent to the cloud, and the average power consumption decreases at a higher rate to

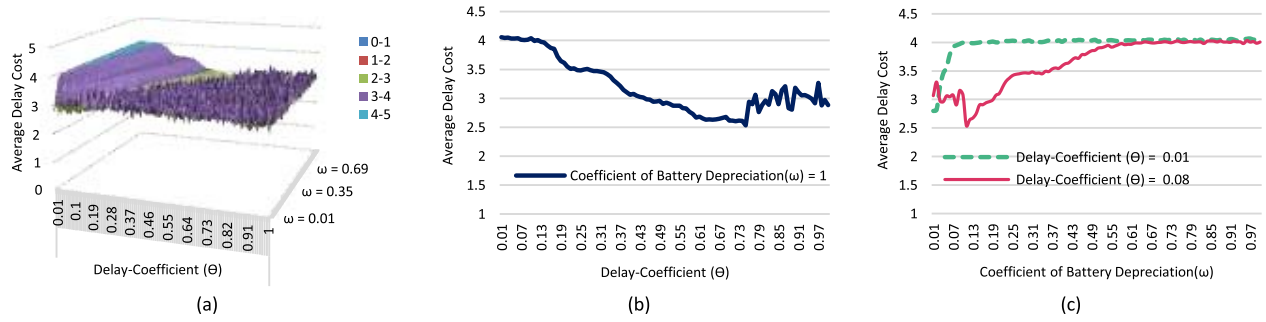


Fig. 4. The average delay cost based on the delay coefficient (θ) and the coefficient of the battery depreciation (ω). (a) The average delay cost by changing the coefficients ω and θ . (b) The effect of delay coefficient (θ) on delay cost. (c) The effect of the coefficient of battery depreciation (ω) on delay cost.

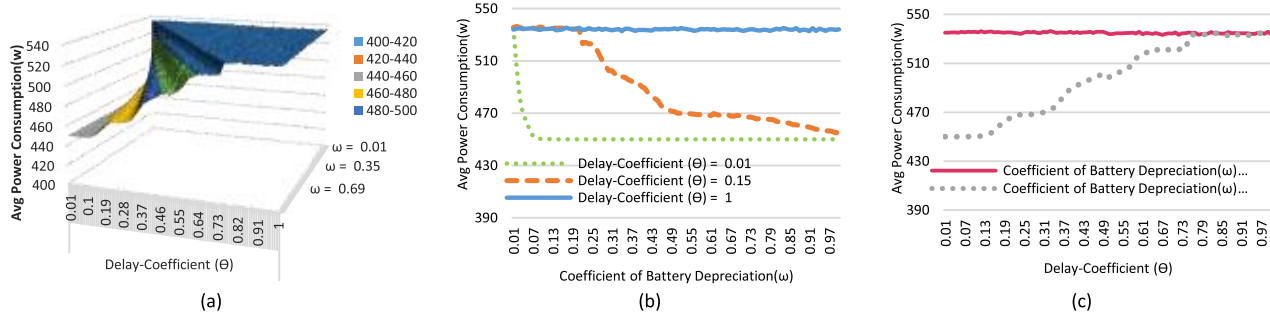


Fig. 5. The average power consumption based on the delay coefficient (θ) and coefficient of battery depreciation (ω). (a) The average power consumption by changing ω and θ . (b) The effect of the coefficient of battery depreciation (ω) on power consumption. (c) The effect of delay coefficient (θ) on power consumption.

achieve its final value (i.e., 450 w). Also, it can be concluded that the rapid decrease in power consumption is due to the minimal effect of delay and the stronger effect of battery power consumption on the fitness function. Another point to mention in this figure is the points on which the delay coefficient $\theta = 1$ is maximum. On these points, due to the strong effect of delay on the cost function, the algorithm sends the majority of processes to the fog server so that they would be conducted locally and the power consumption would not decrease. In this case, the battery level reaches its maximum.

A comparison of the three lines in this graph indicates that the average power consumption of $\theta = 1$ is greater than $\theta = 0.15$ and the average power consumption of $\theta = 0.15$ is greater than $\theta = 0.01$. The high level of power consumption is due to the greater significance of the delay part in the cost function.

Fig. 5(a) shows that, on points with a delay coefficient greater than 0.7, the average power consumption reaches its maximum and remains constant for each state ω . Also, given that $\theta < 0.7$, as the coefficient of battery depreciation increases, attempts are made to send the loads to the cloud and decrease power consumption. As θ decreases, the power consumption part becomes more significant, and the average power consumption is reduced. Fig. 5(c) shows the power consumption graph based on the variations of θ for two values of ω . As can be seen in the figure, when the coefficient of battery depreciation is $\omega = 1$ (maximum), power consumption will increase as the delay coefficient increases and becomes more significant in the cost function. In addition, when the coefficient of battery depreciation is $\omega = 0.01$ (minimum), power consumption will not change with the increase in θ .

This is due to the minimal effect of the coefficient of battery depreciation on the cost function.

Given this, we seek out a state in which the average power consumption is minimized so that the least amount of depreciation could be achieved. As discussed earlier in the formulation of the problem, green energy enters the system as normal distribution according to the equation: $N(520W, 150)$. According to Fig. 5(a), power consumption is almost equal to the average green energy received by the system. It can be thus concluded that this algorithm tries to distribute the workloads in a way that the required power for processing could become almost equal to the green energy and the coefficient of battery depreciation as well as the power consumption at the edge of the network could be minimized. Also, with the decrease in power consumption at the network edge, more green energy could be stored in the batteries.

Fig. 6(a) illustrates the network edge battery levels in different experiments for ω and θ . With the increase in the coefficient of battery depreciation (on points where θ is less than 0.7), more workloads are transmitted to the cloud, which will increase the battery level. Fig. 6 (b) shows the battery level for three constant states of θ . When $\theta = 0.15$ and $\theta = 0.01$, with the increase in the coefficient of battery depreciation (ω), the battery level will increase from 700 w to 2000 w (charging mode). When $\theta = 0.01$ (minimum), due to the transmission of all workloads to the cloud, the green energy not consumed is stored in the batteries and the battery level rises more quickly. However, when $\theta = 1$ (maximum), the loads are maintained at the network edge, thereby leading to the remarkably high power consumption and keeping the battery level at 800 w. It should be mentioned that the proposed algorithm could maintain a full battery in most cases.

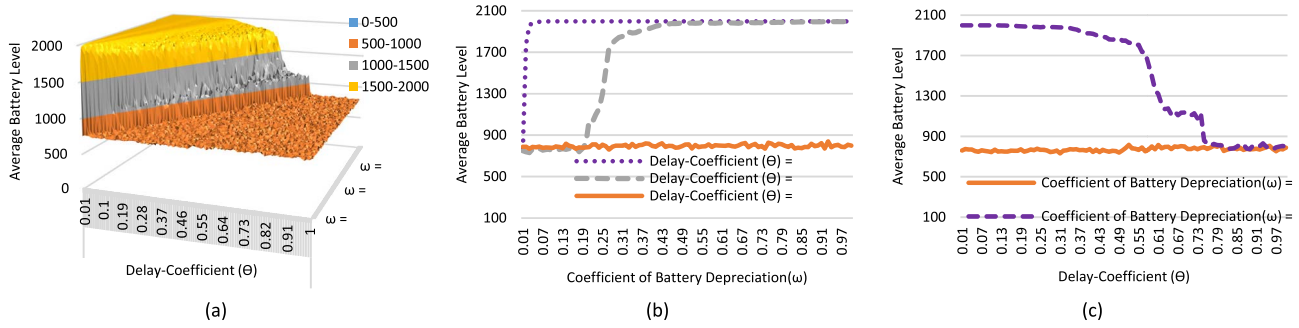


Fig. 6. The average battery level based on the delay coefficient (θ) and the coefficient of battery depreciation (ω). (a) The average battery level by changing ω and θ . (b) The effect of the coefficient of battery depreciation (ω) on battery level. (c) The effect of delay coefficient (θ) on the battery level.

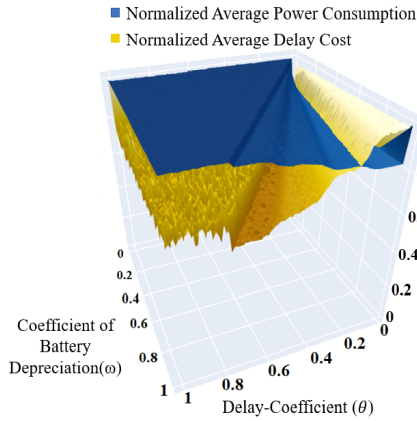


Fig. 7. Normalized values of delay cost and average power consumption.

Comparing the corresponding graphs in Fig. 5 and 6 indicates that, by sending more workloads to the cloud and decreasing power consumption at the network edge, more green energy could be stored in the batteries. This process at the network edge will increase the battery levels. To illustrate this point, let us compare Fig. 5 (c) and 6 (c) in terms of power consumption for the delay coefficient θ and two values of ω . It can be observed that, when the coefficient of battery depreciation is $\omega = 1$ (maximum), power consumption increases with the increase in the delay coefficient as well as its effect on the cost function, thus reducing the average battery level. Also, when the coefficient of battery depreciation is $\omega = 0.01$ (minimum), power consumption will not change with the increase in θ , and the average battery level at the network edge will remain constant. This is not desirable for us because we seek out circumstances in which the average battery level would be maximized. On the other hand, the corresponding graphs in Fig. 4 and 6 indicate that the battery level decreases with the reduction in the delay. The reason is that, in order to reduce the delay costs, the system attempts to process most of the workload in the fog servers, which leads to more battery consumption. Given what was discussed above, we need to reduce the average delay while maintaining the maximum battery level.

C. The Optimum Point of ω and θ

To reach a balance between power consumption and delay in workload distribution, average power consumption and delay cost were normalized to find the optimum state of ω , and θ .

TABLE II
OPTIMUM POINTS BASED ON THE VALUES OF θ AND ω

	1	2	3
Delay Coefficient (θ)	0.05	0.08	0.1
Coefficient of Battery Depreciation (ω)	0.23	0.35	0.47
Average Delay	3.4958	3.49302	3.49707
Average Power	466.59	467.1	466.62

Fig. 7 illustrates the normalized levels of average power consumption and delay cost for every value of ω and θ .

It can be observed that these two parameters have a negative relationship. That is, an increased delay means decreased power consumption and vice versa. As a result, a balance between ω and θ can be attained when the normalized values of delay and power consumption are equal. In other words, the intersection points of the two normalized levels are the points of balance. The intersection of these levels in this figure forms a line. Those values of ω and θ that lie on this line are indicative of a balanced state. Of these points, however, only those points provide an optimum state in which the sum of the two normalized parameters is minimal.

On this basis, the three optimum points from Fig. 7 are described in Table II. This table lists the average power consumption and the delay for each of the points in the parametric space (θ, ω). For a better comparison of the three points, six cross-sectional cuts have been made in the graph in Fig. 7 (Fig. 8 (a) to 8 (f)). In Fig. 8(a), the normalized values for $\theta = 0.05$ can be observed. At the intersection point where the sum of the two parameters is minimal, the cost of battery depreciation should be $\omega = 0.23$. For $\omega = 0.23$, Fig. 8(b) shows that the intersection point at which power and delay are minimal is $\theta = 0.05$. Similarly, for the second optimum point (Fig. 8(c) and 8 (d)), $\theta = 0.08$ and $\omega = 0.35$ achieve a balance, and their sum is the minimum amount. Fig. 8 (e) and 8 (f) depict the third optimum point for ω , and θ . At this point, too, the sum of delay cost and power consumption is minimal with $\theta = 0.1$ and $\omega = 0.47$.

The following are the results of the execution of the proposed workload distribution at the optimum points ($\theta = 0.05$ and $\omega = 0.23$). Fig. 9 illustrates the amount of data processed in the fog, the battery consumption of servers, and the amount of workload sent to the cloud.

In this figure, the ratio of offloading in the cloud and the fog to the total input workload is shown in intervals of 1000. On average, in each interval, 64 percent of the total input load has been processed in fog servers, and the rest

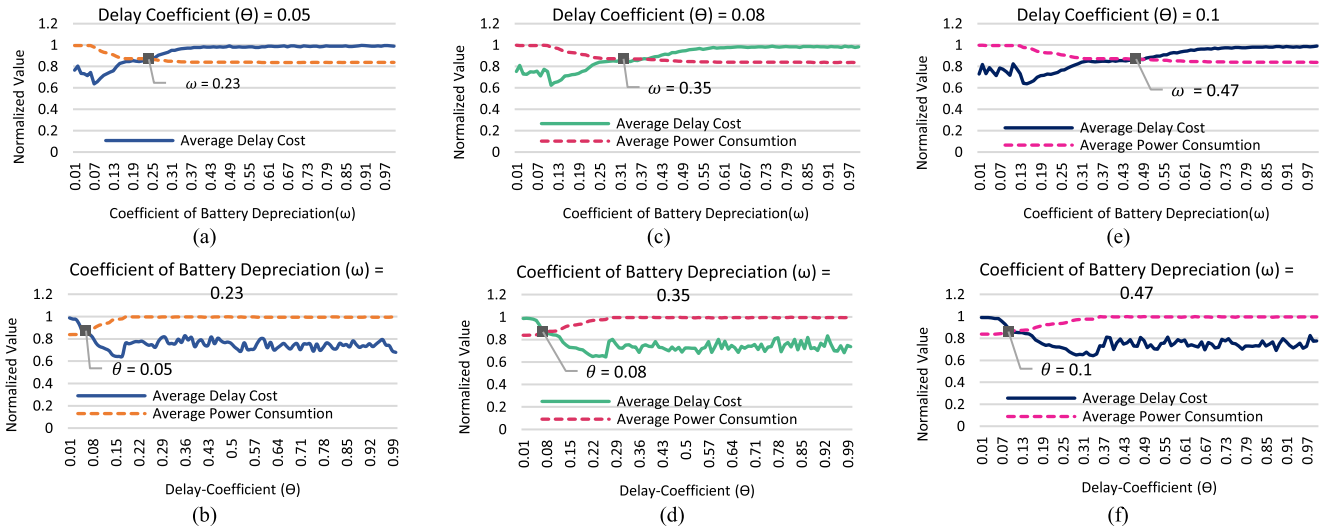


Fig. 8. Normalized values of average delay cost and power consumption for the optimum points. (a) Variations of the coefficient of battery depreciation (ω) for the first point. (b) Variations of delay cost (θ) for the first point. (c) Variations of the coefficient of battery depreciation (ω) for the second point. (d) Variations of delay cost (θ) for the second point. (e) Variations of the coefficient of battery depreciation (ω) for the third point. (f) Variations of delay cost (θ) for the third point.

(around 35 percent) has been sent to the cloud. As most of the loads have been processed locally, it is expected that battery consumption should be high. However, the battery level graph shows that an average of 12 percent of the battery has been consumed in each interval. This can be explained by the optimum use of renewable energy. The system distributes the loads in a way that the power consumed for processing at the network edge be equal to renewable energy. Also, in intervals where more load has been processed locally, there is a rise in battery consumption. For example, battery consumption in the interval 5000-6000 is 3 percent more than in the interval 4000-5000. On the other hand, local processing reduces the delay in the handling workloads.

D. Comparison With Other Methods

In this section, the results of the proposed method at its optimum point are compared with other methods to confirm the decrease achieved in the delay in workload transmission. These methods are briefly described below.

1) *Fixed Power*: In this method, a fixed amount of power is considered for edge computations at each interval of time [24].

2) *Post Decision State (PDS) Algorithm* [20]: The PDS algorithm grabs the state of the system instantly after making the decision at the end of each time interval. The state of the system after making a decision at the end of the interval is an important data that is named the *after-state* variable. The PDS is mainly used as a decision-tree based optimization algorithm. In this algorithm, to find the optimum solution, the problem is broken down into decision nodes and outcome nodes, which correspondingly denote pre-decision and post-decision states. For finding the optimum decision for the vector-valued problem of workload allocation, the PDS tries to find a state that minimizes the long-term costs of the system.

3) *Q-Learning* [25]: Q-learning is considered as a reinforcement learning algorithm that is independent of the type of system model. In this agent-based algorithm, the agent tries to learn a strategy, which results in the best action for each

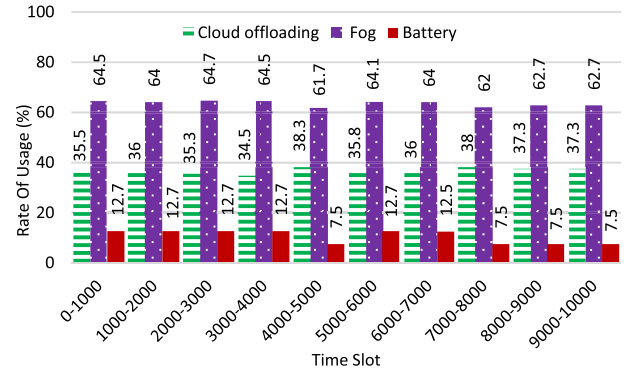


Fig. 9. The rate of usage of cloud and fog resources and power supply over time.

state of the system. Since this algorithm does not need a model of the environment, it can solve the problems with stochastic transitions and payoffs without needing any regulation.

4) *Myopic Optimization* [26]: In this algorithm, regardless of any relationship between the system states and corresponding decisions, the cost function of each state is minimized by only considering the present input information of the system. That is, in the Myopic optimization model, the present knowledge of the workload allocation is densely presented by a Myopic window which represent the knowledge of system in a limited number of time frames. The content of this window may be repeated in different times. As a result, the outcome of the system may be seen repeatedly.

Fig. 10 shows the average delay cost for different methods. As can be observed, learning-based methods perform better and have a lower average delay when run on the battery than when using the electricity network. On the other hand, the proposed algorithm has a lower delay than the other methods. In this figure, the delay cost of all the methods is greater than five, whereas the genetic algorithm used in the proposed method has reduced this cost down to 3.5.

The main point that is clear in both figures is the reduction of the average energy consumption and the reduction of

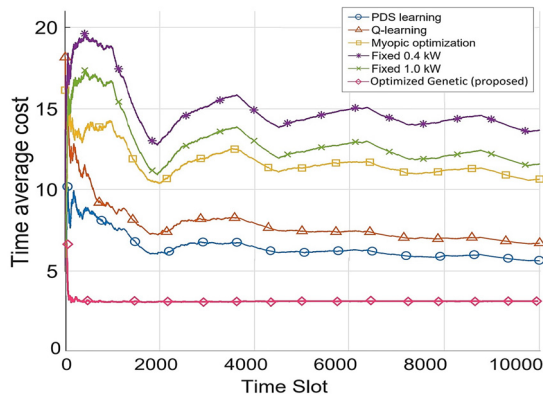


Fig. 10. The average delay cost.

the average delay in successive intervals of time. Reducing the average processing latency for the proposed method in Figure 10 means that workloads are processed more on the fog side, and a smaller percentage of them are sent to the cloud, as evidenced by Figure 9. In Fig.9, for the first 1000 time slots, more percentage of workloads are processed in fog, respectively reducing the average delay. One of the strengths of the proposed method is that it does not have many fluctuations in time slots, especially in the first 2000 time slots.

VI. CONCLUSION

In this paper, we tried to achieve a balance between power consumption at the intelligent vehicular network edge and delay in workload transmission in the clouds by using a genetic algorithm and finding the optimum modes of workload distribution. We also showed that workload distribution at the edge of the vehicular network using renewable energy sources is suitable for vehicular networks in which the processing resources do not have access to the electrical grid and depend on batteries for operation. By utilizing parameters such as the input load and the proportion of green energy as the input parameters of the genetic algorithm, this paper calculated for the first time the optimum number of workloads to be processed locally. Also, by changing the coefficients of the parameters of the cost function of the genetic algorithm, we determined the optimum coefficients for processing the workloads with the least amount of delay and the least power consumption. The simulation results suggest that the proposed method can achieve a better balance in workload distribution than the other existing methods do. While reducing the workload delay by 40 percent and decreasing power consumption at the edge of the vehicular network, this method also seeks to minimize battery consumption by making use of renewable energies.

In future work, other machine learning methods such as neural networks can be used for selecting the optimum parameters.

REFERENCES

[1] Z. E. Ahmed, R. A. Saeed, and A. Mukherjee, "Challenges and opportunities in vehicular cloud computing," in *Cloud Security: Concepts, Methodologies, Tools, and Applications*. Hershey, PA, USA: IGI Global, 2019, pp. 2168–2185.
 [2] T. Islam and M. M. A. Hashem, "A big data management system for providing real time services using fog infrastructure," in *Proc. IEEE Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Apr. 2018, pp. 85–89.

[3] A. Yousefpour *et al.*, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, Sep. 2019.
 [4] M. Shojafar, N. Cordeschi, and E. Baccarelli, "Energy-efficient adaptive resource management for real-time vehicular cloud services," *IEEE Trans. Cloud Comput.*, vol. 7, no. 1, pp. 196–209, Jan. 2019.
 [5] F. S. Abkenar and A. Jamalipour, "EBA: Energy balancing algorithm for fog-IoT networks," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6843–6849, Aug. 2019.
 [6] W. Zhang, Z. Zhang, and H.-C. Chao, "Cooperative fog computing for dealing with big data in the Internet of vehicles: Architecture and hierarchical resource management," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 60–67, Dec. 2017.
 [7] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, "Optimal workload allocation in fog-cloud computing toward balanced delay and power consumption," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1171–1181, Dec. 2016.
 [8] M. Ghobaei-Arani, A. Souri, and A. A. Rahmanian, "Resource management approaches in fog computing: A comprehensive review," *J. Grid Comput.*, vol. 18, no. 1, pp. 1–42, Mar. 2020.
 [9] R. Basir *et al.*, "Fog computing enabling industrial Internet of Things: State-of-the-art and research challenges," *Sensors*, vol. 19, no. 21, p. 4807, Nov. 2019.
 [10] S. Nižetić, N. Djilali, A. Papadopoulos, and J. J. P. C. Rodrigues, "Smart technologies for promotion of energy efficiency, utilization of sustainable resources and waste management," *J. Cleaner Prod.*, vol. 231, pp. 565–591, Sep. 2019.
 [11] M. Aloqaily, A. Boukerche, O. Bouachir, F. Khalid, and S. Jangsher, "An energy trade framework using smart contracts: Overview and challenges," *IEEE Netw.*, vol. 34, no. 4, pp. 119–125, Jul. 2020.
 [12] H. Wu, L. Chen, C. Shen, W. Wen, and J. Xu, "Online geographical load balancing for energy-harvesting mobile edge computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
 [13] Z. Ning, J. Huang, X. Wang, J. J. P. C. Rodrigues, and L. Guo, "Mobile edge computing-enabled Internet of vehicles: Toward energy-efficient scheduling," *IEEE Netw.*, vol. 33, no. 5, pp. 198–205, Sep. 2019.
 [14] X. Wang *et al.*, "Future communications and energy management in the Internet of vehicles: Toward intelligent energy-harvesting," *IEEE Wireless Commun.*, vol. 26, no. 6, pp. 87–93, Dec. 2019.
 [15] H. Chen, T. Zhao, C. Li, and Y. Guo, "Green Internet of vehicles: Architecture, enabling technologies, and applications," *IEEE Access*, vol. 7, pp. 179185–179198, 2019.
 [16] F. Ahmadizar, K. Soltanian, F. AkhlaghianTab, and I. Tsoulos, "Artificial neural network development by means of a novel combination of grammatical evolution and genetic algorithm," *Eng. Appl. Artif. Intell.*, vol. 39, pp. 1–13, Mar. 2015.
 [17] S. Verma, N. Sood, and A. K. Sharma, "Genetic algorithm-based optimized cluster head selection for single and multiple data sinks in heterogeneous wireless sensor network," *Appl. Soft Comput.*, vol. 85, Dec. 2019, Art. no. 105788.
 [18] X. Liu and N. Ansari, "Toward green IoT: Energy solutions and key challenges," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 104–110, Mar. 2019.
 [19] J. Xu and S. Ren, "Online learning for offloading and autoscaling in renewable-powered mobile edge computing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
 [20] J. Xu, L. Chen, and S. Ren, "Online learning for offloading and autoscaling in energy harvesting mobile edge computing," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 3, pp. 361–373, Sep. 2017.
 [21] F. M. Dalvand and K. Zamanifar, "Multi-objective service provisioning in fog: A trade-off between delay and cost using goal programming," in *Proc. 27th Iranian Conf. Electr. Eng. (ICEE)*, Apr. 2019, pp. 2050–2056.
 [22] M. Abbasi, M. Yaghoobikia, M. Rafiee, A. Jolfaei, and M. R. Khosravi, "Efficient resource management and workload allocation in fog-cloud computing paradigm in IoT using learning classifier systems," *Comput. Commun.*, vol. 153, pp. 217–228, Mar. 2020.
 [23] C. N. Giap and D. T. Ha, "Parallel genetic algorithm for minimum dominating set problem," in *Proc. Int. Conf. Comput., Manage. Telecommun. (ComManTel)*, Apr. 2014, pp. 165–169.
 [24] K. Kaur, S. Garg, G. S. Aujla, N. Kumar, J. J. P. C. Rodrigues, and M. Guizani, "Edge computing in the industrial Internet of Things environment: Software-defined-networks-based edge-cloud interplay," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 44–51, Feb. 2018.
 [25] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning* vol. 2. Cambridge, MA, USA: MIT Press, 1998.
 [26] K. Poncelet, E. Delarue, D. Six, and W. D'haeseleer, "Myopic optimization models for simulation of investment decisions in the electric power sector," in *Proc. 13th Int. Conf. Eur. Energy Market (EEM)*, Jun. 2016, pp. 1–9.

Linked Data Processing for Human-in-the-Loop in Cyber–Physical Systems

Zhigao Zheng¹, Member, IEEE, Shahid Mumtaz², Senior Member, IEEE,
 Mohammad R. Khosravi³, and Varun G. Menon⁴, Senior Member, IEEE

Abstract—There are several kinds of smart devices, such as smartphones, sensors, and smart wearable devices, included in the Human-in-the-Loop (HITL) system, but different devices have their own data processing and programming paradigm. Programmers usually need to design the same data processing logic for different devices by using a different programming model. How to mapping the same code to different devices without any change is an emerging topic in the HITL system. Furthermore, the intelligent data processing for the smart CPS sector is experiencing significant growth in data volume, driven by a large number of smart devices that are anticipated in the near future. All these smart devices are expected to improve the overall HITL system performance marvelously. A large number of devices can also outstandingly increase the data volume, which needs to be processed in real time. How to process large-scale data on a smart device in real time is another challenge. Focused on these challenges, this article proposed a computing device-aware HITL CPS data processing framework, named Barge, aiming to map the regular code to the different hardware without any change. In Barge, a semantic model, an architecture-driven programming model, and a graph partition scheme are included. The semantic model is used to express the user-defined graph algorithms by using the domain-specific language. The architecture-driven programming model will execute the graph algorithms on a different device in parallel. Furthermore, the graph partition scheme will partition the large-scale graphs into suitable partitions by aware of the topology to make the partitioned data suitable for kinds of smart devices. We believe that our work would open a wide range of opportunities to improve the performance of large-scale graph processing for HITL systems.

Index Terms—Cyber–physical systems (CPSs), data partition, graph computing, Human-in-the-Loop (HITL), new architecture, programming model, semantic model.

I. INTRODUCTION

THE tremendous amount of smart devices, such as smartphones, radio frequency identification (RFID), sensors,

Manuscript received March 30, 2020; revised June 30, 2020 and August 19, 2020; accepted September 9, 2020. Date of publication April 9, 2021; date of current version September 30, 2021. (Corresponding author: Zhigao Zheng.)

Zhigao Zheng is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: zhengzhigao@hust.edu.cn).

Shahid Mumtaz is with the Instituto de Telecomunicações, Universitário de Santiago, P-3810-193 Aveiro, Portugal (e-mail: smumtaz@av.it.pt).

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 71557-13876, Iran (e-mail: m.r.khosravi.taut@gmail.com).

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

Digital Object Identifier 10.1109/TCSS.2020.3029569

and embedded devices, have revolutionized both the physical and digital world through the integrated interactions to create the global smart cyber–physical systems (CPSs) [1]–[3]. To process the large scale of the complex linked data between different kinds of devices, both data-intensive and memory-intensive data processing frameworks are included in CPS, i.e., CPS is a software-intensive decentralized system that autonomously perceives its operational context [4], [5]. A CPS system consists of the hardware infrastructure (physical components) and software model [6]–[8]. The most important software is the cyber twin, which is used to simulate the physical things. Internet of Things (IoT), on the other hand, connected kinds of sensors and some other physical things. This means that IoT acts as a connection bridge to network different cyber–physical things. CPS, also known as big data processing technologies, is a hot topic, which leads to a set of new research interests, and it was widely used in many services, such as customer behavior prediction and weather and environment monitoring. However, most of the data processing logic for the same application is the same, but programmers need to develop multiple different copies of code for an application and deploy them on different devices. How to release programmers from the strenuous repetitive work is an emerging topic in the HITL CPS system. Furthermore, all these services will generate an enormous amount of data in real time, which makes it is not easy to store and process such kind of large-scale data. All these difficulties drive the scientist to propose cloud computing technologies along with machine tools, data mining, artificial intelligence, and fog computing technologies to store, process, and analyze large-scale data. By using all these technologies, we can try to uncover the hidden patterns, unknown correlations, and other useful information [9], [10]. The characteristics of big data were well summarized in the Introduction Section of [11]. The relevance of the big data era and CPS are also highly relevant to global sustainable development goals recently discussed in [12].

Graph computing is one of the most famous big data processing technology, which was widely used to process the linked data. In the graph computing paradigm, the graph data model, which is a fundamental mathematical structure used to model pairwise relations between objects, is widely used in machine learning [13], [14] and deep learning [15] technologies to express the connections between different objects. The context in a graph is called vertices (also called nodes), while the links are called edges. Graph theory has

been widely used in Human-in-the-Loop (HITL) data processing [16]–[18], a knowledge graph programming with an HITL discussed in [16]. In Lou’s work [16], the authors examined the advantages of the knowledge graph programming for HITL, such as the flexible programming interface and kinds of “data compiler” method. Then, the authors proposed a knowledge graph programming prototype for HITL. Holzinger *et al.* [17] provided new experimental insights on how to improve computational intelligence by complementing HITL with human intelligence in an interactive machine learning approach. The article [18] described a “big picture” of HITL data analysis, including the user communities’ tools and algorithms, and also the HITL data analysis framework developing technologies and theories. However, how to use graph theory and algorithms to support distinctive characteristics of HITL CPS data analysis and provide high-performance and real-time decision-making policy remains challenging and represents a promising research direction.

With the development of hardware manufacturing, there are several kinds of new computing devices proposed for large-scale data processing, and traditional large-scale graph computing is facing new opportunities and challenges. Graph applications have poor locality and poor cache hit rate and are largely stalled on memory accesses since there are complex connections between graph nodes and the working set of realistic graphs is much larger than the last level cache (LLC) of current machines. The conventional computing architecture is computing-centric, which focuses on memory sharing and message communications; this processing fashion is unable to handle graph applications. In this article, we focus on the key technologies and methods of a new computing architecture for large-scale HITL CPS graph data processing. To solve the poor locality and poor cache hit rate problem, we proposed new computing device-based HITL CPS data processing architecture, named Barge. We made the following contributions to the proposed architecture.

- 1) We conduct an extensive set of comprehensive experiments to explore the parallelism and memory operations of the graph processing systems for HITL CPS data processing.
- 2) We proposed a semantic model for large-scale graph data processing under new computing architecture to mapping the same code to different devices without any change. The semantic model includes semantic rule and graph data interpretation method.
- 3) We proposed an architecture-driven programming model, which is suitable for a different architecture.
- 4) We proposed a data- and topology-aware graph data partition scheme that can partition the large-scale graph data quickly by consider the data structure and also the feature of the computing device.

The rest of this article is organized as follows. We will introduce the characteristics and research challenges of HITL CPS data processing and the motive of using the graph processing method to process the large-scale HITL CPS data in Section II. Section III provides our proposal for applying the proposed Barge model to improve the performance of HITL CPS data processing. Then, we will introduce the design methodologies

and principles of Barge in Section IV. Section V introduces the related work, and we will conclude this article in Section VI.

II. CHALLENGES OF HITL CPS DATA PROCESSING

In this section, we describe our experimental settings, including the graph data sets and algorithms and the graph processing frameworks (GPFs) for our empirical study. We also try to explore the parallelism issues and memory operation characteristics of GPFs for HITL CPS data processing.

A. Experimental Settings

1) *Graph Algorithms*: Most of the graph algorithms can be classified as the iterative algorithm and the traversal algorithm. All vertices will be updated in each iteration of the iterative algorithm, but only active vertices will be updated in each iteration of the traversal algorithm. The most representative algorithms of iterative and traversal algorithms are PageRank (PR) and Breadth-First Search (BFS), respectively. We select these two most representative graph algorithms for the performance evaluation, as well as much prior work [19], [20] do.

- 1) *BFS*: An algorithm that traverses the whole graph in search of one or more vertices, which is a basic component of many other complex graph mining algorithms.
- 2) *PR*: An algorithm that was used to evaluate the influence of vertex within a graph, which was proposed by Google first, and it was initially used to evaluate the importance of a web page.

2) *Graph Processing Frameworks*: Many GPFs have been developed by both academic and industry researchers, such as TOTEM [21], CuSha [22], and some other frameworks. In this section, we select four representative state-of-the-art GPFs to run the graph algorithms and profiling the runtime details.

- 1) *GunRock [19]*: A high-performance graph processing library on GPU with a high-level bulk-synchronous processing scheme. Gunrock provides a data-centric abstraction centered on operations on a vertex or edge frontier. Gunrock achieves a balance between performance and expressiveness by coupling high-performance GPU computing primitives and optimization strategies. Gunrock also proposed a high-level programming model that allows programmers to quickly develop new graph primitives with small code size and minimal GPU programming knowledge.
- 2) *Sep-Graph [20]*: A highly efficient software framework for graph processing on GPU. It provides a hybrid execution mode that can automatically switch among synchronous or asynchronous execution mode, Push or Pull communication mechanism (Push or Pull), and Data-driven or Topology-driven traversing scheme (DD or TD), according to the parameters.
- 3) *Tigr [23]*: A graph transformation framework that can effectively reduce the irregularity of real-world graphs with correctness guarantees for a wide range of graph analytics.

TABLE I
DATA SETS USED IN THE EXPERIMENTS

Datasets	Vertices	Edges	Avg. Degree	Max Degree	Diameter	Exponent(α)	x_{min}	Fitness (p)
web-Stanford	281,904	2,312,497	8.20	38,626	164	2.1310	5	0.894
dblp-2011	986,208	6,707,236	6.80	979	23	3.9736	119	0.4
youtube	1,157,829	2,987,624	5.27	28,754	24	2.1410	8	0.877
RoadNet-CA	1,971,282	5,533,214	2.81	12	8,440	15.5587	4	0
Wiki-Talk	2,394,386	5,021,410	4.19	100,032	11	2.4610	1	0.787
soc-LiveJournal	4,847,571	86,220,856	28.25	22,887	20	2.6510	59	0.930
twitter-2010	41,652,230	1,468,365,182	70.51	3,081,112	23	1.54	12	0.96
webbase-2001	118,142,155	1,019,903,190	8.63	816,127	379	2.2	6	0.538

4) *GSWITCH* [24]: A machine learning model-based graph processing system that dynamically chosen the optimization variants by monitoring the system overhead. In *GSWITCH*, the authors trained 644 real graphs to learn the algorithm pattern, and *GSWITCH* changes the optimization variants by using the pattern information to achieve high parallel system performance.

3) *Graph Data Sets*: All the data sets of the experiments conducts in this article are follow the classic graph formalism [25]. We use V and E to present the vertices and edges of the graph, respectively. $G = (V, E)$ present the graph. The edge presented as e , where $e = (u, v)$ and $e = \langle u, v \rangle$ are the undirected and directed edges. In this article, both directed and undirected graphs are used in our experiments. In order to show the performance of different kinds of graphs, we include both power-law and large diameter graphs in all our experiments. We select eight graphs from different real-world applications, such as e-business, social network, and some other source networks, and all these graphs are with different structures and a varying number of vertices and edges. The graphs are shown in Table I. All these eight graphs can be downloaded from the Stanford Network Analysis Project (SNAP) [26]. As the power-law graph follow a distribution, as shown in formula (1) [27], [28], we list the exponent and the fitness in Table I to compare the power-law attribution. In this article, the graphs are stored in a plain text file with an edge-list format, which is easy for us to locate the edges

$$\mathbb{P}(x) \propto x^{-\alpha}. \quad (1)$$

4) *Hardware Platforms*: All the experiments presented in this section are conducted on NVIDIA Tesla V100 GPU, which is a Volta architecture-based GPU with 5120 CUDA cores and 32-GB onboard memory. The GPU is coupled with a host machine equipped with 28-Ksyun Virtual Senior CPU cores, each at 2.60 GHz, and 12-GB memory. The host machine is running Ubuntu OS version 16.04.10. The algorithm is implemented with C++ and CUDA 10.02 using the “-arch=sm35” compute compatibility flag.

B. Exposing More Parallelism

To improve the parallelism of hardware, the Single Instruction Multiple Data (SIMD) architecture is introduced to the out-of-order (OOO) manner to enable simultaneous execution of multiple independent instructions. In this design philosophy, each core could issue more than one Micro-Operations (μ -op).

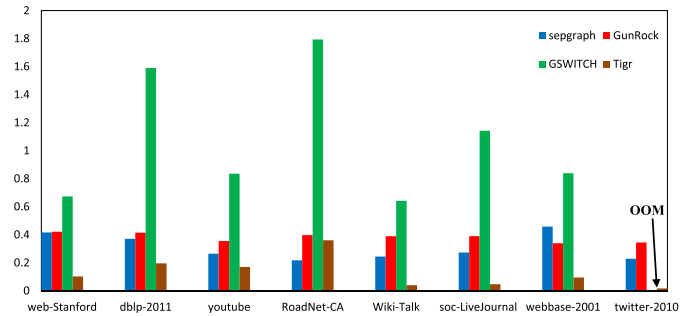


Fig. 1. Average IPC of PR on different data sets with different implementations.

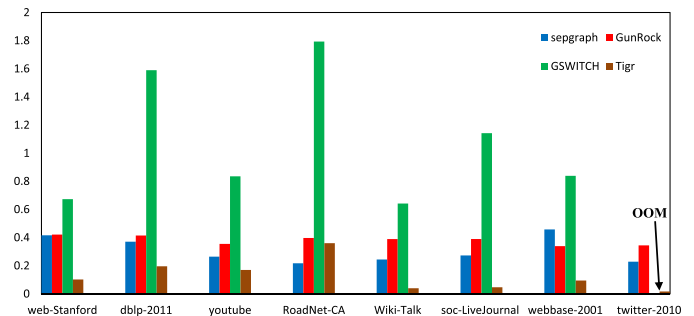


Fig. 2. Average IPC of BFS on different data sets with different implementations.

We illustrate the instruction-level parallelism (ILP) of all the evaluated frameworks. Figs. 1 and 2 show that the average instructions per cycle (IPC) of PR on *GSWITCH* is about 1.793, while the IPC of PR on the other three frameworks is no more than 0.458. This experiment indicates that only about one-eighth of the core’s ability is used for most existed GPFs (except *GSWITCH*). One main reason for low IPC is the long instruction latency [29], and there is a large number of GPU cycles consumed by some instructions.

In order to identify the exact reason for the low ILP phenomena, we also evaluated the Max/Min IPC of different implementations of PR and BFS on different data sets. Figs. 3 and 4 show that the Max IPC of PR and BFS on Sepgraph, Tigr, and GunRock are no more than 0.543, while the MAX IPC of PR on *GSWITCH* is 1.823. This phenomenon indicates the heavy dependence on graph processing algorithms. The execution of one instruction may relate to many other instructions.

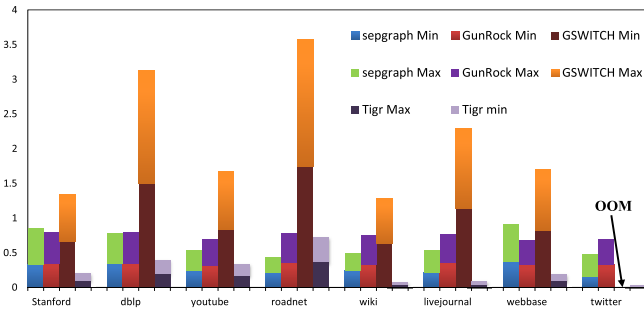


Fig. 3. Max/Min IPC of PR on different data sets with different implementations.

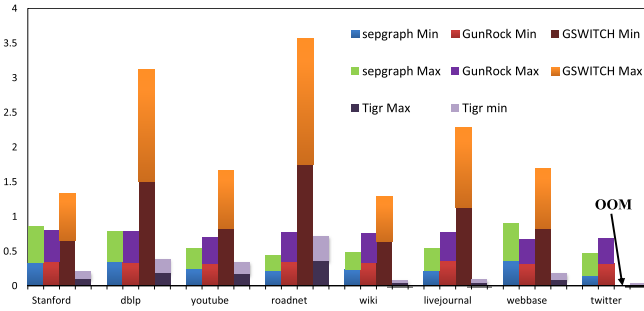


Fig. 4. Max/Min IPC of BFS on different data sets with different implementations.

C. Warp Issue Efficiency

To look deeper into the reason for low ILP, we evaluated the occupancy of GPU warps. In GPU performance profiling, the achieved occupancy is used to measure the warp scheduler’s efficiency by using the ratio of the average active warps of each clock cycle to the maximum warps supported by each multiprocessor (SM), which defined as for the following formula:

$$\text{occupancy} = \frac{\text{active warps}}{\text{maximum warps}}. \quad (2)$$

From formula (2), we can conclude that low occupancy interferes with the ability to hide memory latency, which can degrade the performance. In contrast, higher occupancy does not always indicate higher performance. Many factors can affect occupancy, such as register availability. The register storage differs for different devices, and it enables the threads to keep local variables nearby for low-latency access. However, the threads must share the register file due to a limited commodity. In modern GPU architecture, all the registers are allocated to a block at the beginning of the program. Hence, a supported block of a multiprocessor (SM) will be reduced if each block uses more than one register, and this thread assignment will further reduce the occupancy of the SM.

Figs. 5 and 6 show the occupancy is very low of both PR and BFS implementation of GunRock on all the eight graph data sets, while both the iterative and traversal algorithm can achieve high occupancy of Tigr and GSWITCH implementation.

D. Memory Operations

In this section, we explore the memory operation efficiency by checking the global memory efficiency and throughput.

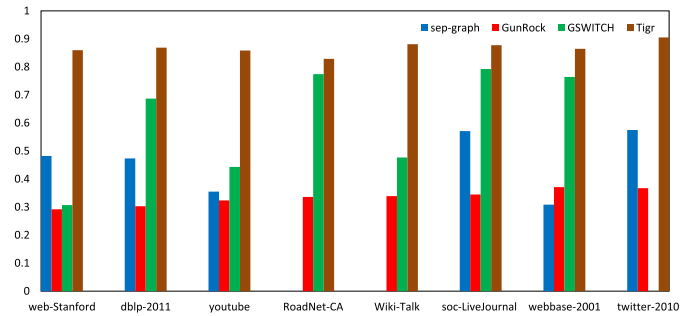


Fig. 5. Achieved occupancy of PR on different data sets with different implementations.

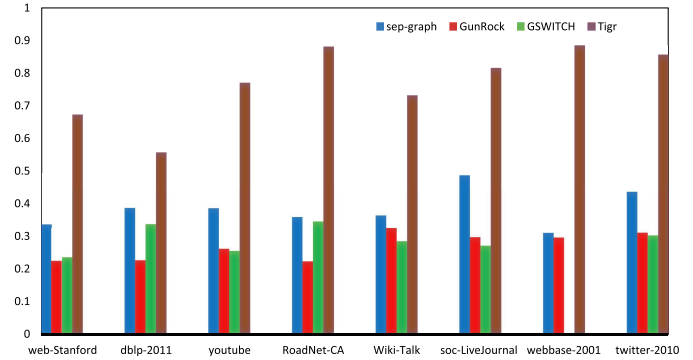


Fig. 6. Achieved occupancy of BFS on different data sets with different implementations.

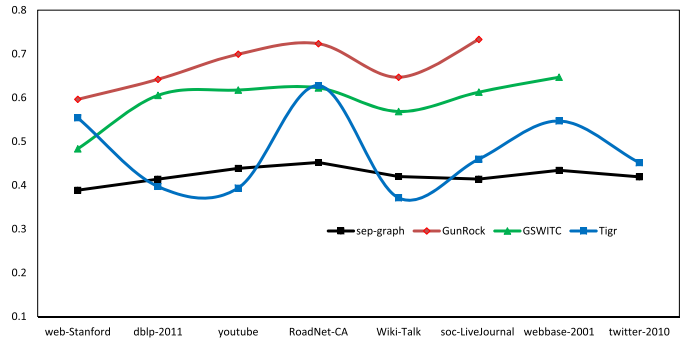


Fig. 7. Global memory efficiency for BFS on different data sets with different implementations.

Figs. 7 and 8 show the global memory efficiency for BFS and PR on different data sets with different implementations. We can conclude from Figs. 7 and 8 that GunRock can achieve the highest memory operation efficiency for both the traversal and iterative algorithms while Tigr and GSWITCH tending to vary widely on different data sets.

Table II shows the global memory throughput for BFS and PR on different data sets, and Tigr can achieve the highest global memory throughput on both BFS and PR. GunRock and GSWITCH achieve the lowest throughput on most power-law graphs, while Sep-Graph can achieve higher throughput on the large-scale graphs (RoadNet-CA) than the power-law graphs.

III. APPLYING BARGE IN GRAPH-BASED HITL CPS DATA PROCESSING

As we discussed in Section II, there are several challenges for graph processing on GPU, such as the control and memory

TABLE II
GLOBAL MEMORY THROUGHPUT FOR BFS AND PR ON DIFFERENT DATA SETS

Dataset	Algorithm	sep-graph	GunRock	GSWITCH	Tigr
web-Stanford	BFS	129.282	20.455	46.950	574.200
	PR	239.054	218.939	141.676	385.350
dblp-2011	BFS	211.695	31.467	57.259	303.110
	PR	317.688	230.4609	180.500	650.760
youtube	BFS	127.786	105.280	67.061	388.960
	PR	86.187	224.371	314.670	519.400
RoadNet-CA	BFS	234.096	11.830	21.154	632.380
	PR	NULL	285.526	446.260	876.900
Wiki-Talk	BFS	128.983	210.178	93.004	387.250
	PR	NULL	255.578	217.316	82.537
soc-LiveJournal	BFS	195.8851	154.3607	107.335	692.410
	PR	189.973	290.041	377.910	143.770
webbase-2001	BFS	276.6052	221.7952	NULL	532.160
	PR	318.656	313.666	210.889	383.780
twitter-2010	BFS	237.222	173.589	197.400	820.510
	PR	181.970	325.354	NULL	48.873

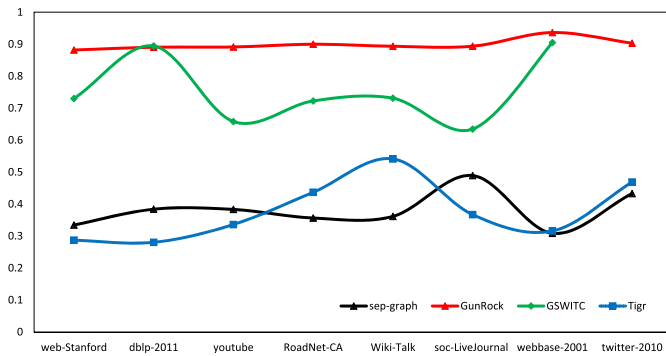


Fig. 8. Global memory efficiency for PR on different data sets with different implementations.

divergence, load imbalance, and global memory access overhead [30]. To achieve an efficient performance of large-scale HITL CPS graph data on new architecture devices, this section introduces the Barge framework for HITL CPS graph data processing.

We design Barge by considering the three primary aspects of graph processing: algorithm semantic expression, programming model, graph partition, and data placement.

- 1) Existing graph processing systems are designed for single hardware by considering the hardware feature to improve the system performance. While this design philosophy cannot achieve expected performance on new hardware architecture, in some cases, the existed frameworks even cannot run on the new device. This design philosophy is easy to limit the scalability of the framework and will lead to strenuous and repetitive work for programmers. In order to solve this problem, this article proposed a semantic model to represent the graph algorithm and make the graph algorithm suitable for the new architecture without reprogram the algorithm. The semantic model provides a set of unified semantics to define the graph structure and a set of unified API for different graph processing systems.
- 2) In order to remove the conflicts between the heterogeneous parallelization of kinds of new hardware devices

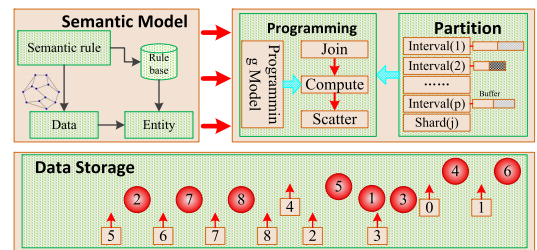


Fig. 9. Illustration of the proposed HITL CPS data processing framework.

and the scalability requirements of graph processing, this article proposes an architecture-aware programming model that can be suitable for multiarchitecture. The proposed model provides a reasonable run-time abstraction of hardware parallel features and the rich programming interfaces for graph computing applications.

- 3) The new architectures devices usually have very high local memory access bandwidth but are just equipped with a small capacity of onboard memory. How to use the limited memory of new architectures devices to process a large-scale graph is a great challenge. In order to solve this problem, this article proposes a new graph partition scheme by considering how to balance the computing workload and communication overhead and the memory access efficiency and also the redundant computation overhead.

Fig. 9 shows an illustration of the proposed HITL CPS data processing framework. In this framework, we propose a semantic model to represent the graph algorithms, a new programming model to fit multiarchitectures, and also a new graph partition scheme.

A. Graph Algorithm Semantic Model for New Architecture

Nowadays, with the proliferation of mobile devices and wearable devices, the HITL CPS graph data are proliferating. The efficient parallelism graph processing algorithm is the most essential aspect to improve the performance of graph processing systems. Combining the features of the

new architecture with the characteristics of graph processing algorithms is the key to improving the performance of graph processing systems. Traditional work usually makes different implementations for each architecture. This method has poor portability and does not meet the needs of multiple graph processing algorithms. It also brings new challenges to implement the algorithms for different kinds of applications and also code management. However, some common operations, similar optimization techniques, and even the same software components are included in various graph algorithms. Considering the abovementioned architecture features and graph processing algorithm characteristics, this article proposes a new graph algorithm description semantic model. This model provides users with productive semantics operations, such as to define graph data structures, describe architecture-aware parallel operations, and fit for different graph algorithms.

The semantic model is an abstraction of the common operations of different algorithms; hence, how to extract the ordinary operations of different algorithms, define the specific operations, and provide a set semantics to describe the parallelizable operations are the base of the semantic model. On the other hand, the semantic model should clearly define parallel operation rules to ensure that user code can be executed correctly and efficiently under the new architecture. At the same time, the semantic model should provide productive parallel operation methods to ensure that users can obtain sufficient expression ability to intuitively, accurately, and completely and describe existing graph algorithms and graph algorithm flows that may appear in the future. How to ensure the efficiency of the graph algorithm by considering the characteristics of the new architecture is the main content of the semantic model. Different architectures, such as GPUs, FPGAs, the SIMD acceleration components, and specialized devices with complicated local storage (such as IBMCELL/Intel SCC), have their parallel execution models. The difference in the execution model of different devices can lead to substantial performance gaps for the same code. In order to achieve excellent overall system performance on different devices, this article first studies how to ensure that the graph algorithm description, which provided by the user, can run on different devices. Furthermore, this article introduced how to determine a parallel operation execution mode to adapt the execution characteristics of the hardware.

B. Architecture-Aware Graph Computing Programming Model

The graph computing programming model is a bridge between the architecture and the graph processing application. On the one hand, it abstracts the details of the hardware characteristics and provides programmers with rich interfaces to implement various graph algorithms. On the other hand, it makes full use of hardware resources efficiently and correctly completes the application requirements, which provides an optimized run-time environment for programs. Graph processing is a strong data dependence application that is hard to parallelism, while most of the new architecture devices are highly parallelism. Hence, how to concisely implement

various graph algorithms while ensuring the efficiency of parallel computing is the first problem to be studied in graph computing programming models.

Although various new-type processors are highly parallelism, their structures are very different. For example, the hardware logic of the computing core of GPU is straightforward, which is suitable for sequence programs without complex logic, while KNL has fewer computing cores than GPU, but the processing logic is relatively complex, which can support more instructions, such as AVX512ER and AVX512CD. Furthermore, the FPGA achieve hardware-level programming by changing the state and combination of gate circuits, which supports some complex logic with high memory bandwidth. These heterogeneous parallelisms make it challenging to achieve excellent system performance in a unified programming manner. Therefore, how to design a run-time optimization mechanism by considering the hardware characteristics, which can be suitable for multiple architectures, is another research challenge for HITL CPS graph processing.

In order to solve the abovementioned two problems, this article proposes an architecture-aware graph computing programming model to match the different architectures. The proposed programming model can effectively and concisely implement various graph algorithms by considering the device characteristics. We can also provide a set of efficient programming methods for developers by using this programming model.

C. Feature-Aware Data Partitioning and Placement Strategy

New accelerator architecture-type processors have large on-chip memory bandwidth, such as GPU and MIC's HBM, which provides very favorable processing conditions for memory-intensive graph computing applications. However, the scale of the HITL CPS graph data shows an explosive growth trend. The growth of the on-chip storage of processors with accelerated structure types is far from meeting the demand for data growth. At the same time, graph processing has some unique features, such as the most graph applications that can be processed by using the iteration processing fashion; on the other hand, the data placement also has a huge impact on the performance of graph algorithms. How to design a suitable data placement strategy and an efficient data partitioning scheme, by considering both the features of the graphs and also the new architecture devices, allocate the partitioned graph data are the crucial points to improve the performance of HITL CPS GPF.

Allocate the partitioned graph data for both single and multimachine, which will cause a large amount of network overhead and bus data transmission overhead on new architecture devices, due to the association of the graph data. It will lead to some other problems, such as load imbalance and low resource utilization, if we only focus on communication and transmission overheads. A high-quality partitioning algorithm itself will cause huge overhead; hence, it is essential for us to study how to reduce the computational complexity of the partitioning algorithm with acceptable partitioning quality.

The graph diameter becomes more extensive due to the rapid growth of the graph size, which leads to lots of redundant

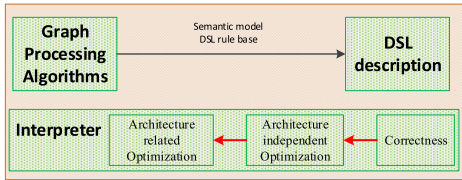


Fig. 10. Graph computing semantic model adapted to new architecture.

computations under the iterative execution fashion on a new architecture device. On the one hand, in the edge centric processing model, the widely used CSR/CSC format easily leads to scattered and irregular memory access, as well as the low memory bandwidth efficiency. On the other hand, the power-law characteristics will lead to a load imbalance problem. In order to solve these problems caused by the data format, this article needs to consider how to improve reading efficiency and how to reduce reading times. In detail, one is how to read the necessary data from the global memory quickly, and the other one is how to implement a heuristic method to read the data, which will be executed in the next iteration, in one memory access to reduce the I/O overhead.

IV. GRAPH-BASED HITL CPS DATA PROCESSING FRAMEWORK

In this section, we first analyze the characteristics of different computing devices and then design a graph algorithm semantic model for these new architectures, and we also abstract the parallel operations for all these architectures in this section. Furthermore, we summarize the data access of large-scale graph processing pattern, and then, we design a multiarchitecture supported graph computing programming model to improve the system performance of large-scale graph processing on new architectures, by simplifying the semantics of the graph processing programming model. Finally, we design a structure-aware data partitioning and placement strategy to make the graph processing system meet the architectures' feature.

A. Graph Algorithm Semantic Model for New Architecture

This article proposes a semantic model of graph algorithms on new architectures by using a domain-specific language (DSL).

The semantic model uses DSL as its entity and uses DSL to express the model elements, such as variable definitions, data definitions, operation definitions, and parallelization flags required in the semantic model. Based on DSL, this article proposes an interpreter for the semantic model. The DSL provides users with a clear and intuitive description of graph algorithms, and the semantic model interpreter explains the description of user-provided graph algorithms. The proposed semantic model is shown in Fig. 10.

This article will analyze the graph processing from a mathematical perspective and then design the semantic model for graph algorithms. A graph $G(V, E)$ is a set of vertices V and an edge set E . The edge and vertex related data can be defined

as a mapping P , P maps the vertex or edge to a particular domain R , and the mapping can be represented as $P : E \rightarrow R$ or $P : V \rightarrow R$. This article uses the mapping to represent the attributes of the edges or vertices. For a given graph $G = (V, E)$ and a series of attributes $\Pi = P_1, P_2, \dots, P_n$, the proposed semantic model should satisfy the following types of graph algorithms: 1) calculate a scalar value from a given graph G and the attribute set, such as the calculating the conductivity of the subgraph; 2) calculate the new attribute from Π , such as calculating the PR value to sort the vertices of the PR algorithm; and 3) select the interested subgraphs, such as the strongly connected components finding algorithm.

Furthermore, the proposed semantic model tries to provide three ways to describe parallel algorithms. The three description ways are as follows: 1) implicit parallelism semantic structure; 2) allow users to distinguish parallel regions accurately; and 3) Well-defined parallel operations. For example, the *for-each* statement is precisely a parallel execution area specified by the user. At the same time, the graph vertex value assignment is an implicit parallel operation, and the widely used reduction operation in graph algorithms is an explicit parallel operation.

This article designs the semantic model interpreter from the following three aspects.

- 1) Check whether the description of the graph algorithm by the user meets the model's requirement. The basic condition is that the user's description should meet the semantic model's grammatical requirements. The interpreter checks the user's input by checking the syntax, data type, and parallel semantics three aspects.
- 2) *Architecture Independent Optimization*: The interpreter converts the code that meets the syntax requirements into a detailed loop or iterative operation and then optimizes the code by cyclic fusion, statement upward and slack protocol boundaries operations.
- 3) *Architecture-Related Optimization*: As a different architecture has its execution fashion and data access method, some architecture-related optimizations should be added through the interpreter by considering the architecture's feature. For example, the GPU uses SIMD fashion to execute the codes in parallel, and the continuous memory access operation can reduce the memory access overhead. Therefore, some data-level parallelism and memory accessing optimization methods should be added through the interpreter. All these optimization methods have a strong correlation with the programming model. Hence, we will introduce them in Section IV-B.

B. JCS Programming Model

Most of the existed GPFs adopt the vertex-centric programming model since this programming model is easy to express most of the graph algorithms, and it can provide high scalability of the graph algorithms by partition the graphs. However, this programming model is also easy to lead random memory access and load imbalance problem due to the skewed degree distribution of real-world graphs. While the edge-centric programming model will introduce

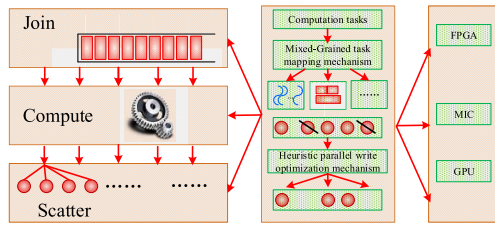


Fig. 11. Graph computing programming model adapted to multiple hardware characteristics.

lots of redundant computation because there are many more edges than the vertices in real-world graphs, it can provide a continuous memory access fashion. The Gather–Apply–Scatter (GAS) programming model proposed in PowerGraph [31] is a fine-grained vertex-centric programming model. In GAS, the computation process is subdivided to increase the degree of parallelism. Previous research shows that there are a number of active vertices in each iteration, which is far less than the total number of vertices in a graph [31]. Hence, this article proposes a queue-based vertex-centric Join–Compute–Scatter (JCS) programming model, which is shown in Fig. 11. In the JCS programming model, the operation on the vertex is divided into the join, compute, and scatter three steps. The join operation adds the active vertices into the worklist, and the compute operation updates the vertex’s value according to the user-defined function, while the scatter operation scatters the vertex’s value to its neighbors, just similar to the scatter operation in the GAS model. In the JCS model, each iteration cares about the vertices that need to be updated. This execution fashion can provide a unified and concise implementation fashion for different algorithms, and it can provide high scalability for the vertex-centric method.

In order to make three phases of the JCS model suitable for different hardware, this article provides a mixed granularity task mapping mechanism and a heuristic parallel optimization mechanism. We introduce the two optimization mechanisms as follows.

- 1) **Mixed-Granularity Task Mapping Mechanism:** Here, we take GPU as an example to introduce the mixed-granularity task mapping mechanism. We assign the vertices to thread, warp, CTA, and kernel according to the number of the neighbor of the active vertex, and the virtual-warp is used to solve the load imbalance problem. While, on KNL, the proposed mechanism not only supports regular round-level parallelism for different task size, including the *for-all* parallelism, reduction parallelism, and scan parallelism, it also supports the irregular parallelism, which can achieve excellent load balancing through reasonable task stealing scheduling.
- 2) **Heuristic Parallel Writes Optimization Mechanism:** Since there are some duplicate vertices in the worklist and some vertices connected with the same vertex, this situation will lead to conflicts in updating operation. Atomic operations and locks are used to ensure data consistency. However, many existed researches show that some updating operations are idempotent (i.e., the updating order does not affect data consistency).

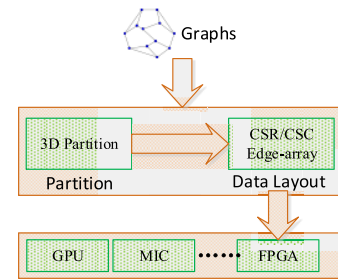


Fig. 12. Data partitioning and placement strategies for data-aware and structure-aware.

Hence, there is no need to design a specific algorithm by using atomic operation or lock to remove the duplicate vertices from the worklist, and just a runtime lightweight heuristic method is enough to remove the redundant computation, which will be more efficient.

C. Feature-Aware Data Partitioning and Placement Strategy

In order to meet the architecture’s feature to unleash the device performance, this article proposes a mixed 3-D graph partition scheme. The proposed graph partition scheme will load the graph blocks into the device on broad memory to make sure the locality of data access and, hence, to reduce the I/O overhead by considering the communication overhead. Fig. 12 shows the essential operation of the proposed graph partition scheme.

The existed graph processing system applied the 1-D or 2-D partition strategy, which is the vertex-centric and edge-centric partition method, respectively. The vertex-centric partition strategy will lead to a load imbalance problem, since the vertex degree distribution of most real world graphs is power-law. While the edge-centric partition strategy will lead to a large amount of communication between the master node and the replicas. Recent research proposed a 3-D partition strategy, which partitions the vertex attribution as the 3-D partition object, and this partition can achieve an excellent system performance in machine learning applications but cannot suitable for full broad applications [32]. While the traversal tree-based partition method can maintain good locality, the partition operation can be executed after traverse the whole graph, and the renumber operation is needed for the subgraphs. The overhead of this partitioning method is huge, and the overhead will increase growth with the graph size. In order to solve the problems of the existed partition methods, this article proposes a hybrid partition method. The hybrid partition strategy will partition the vertices that have similar degrees together by using a 2-D partition method and then partition the subgraphs again by using a 1-D partition method. While, for some specific graph algorithms, the hybrid partition strategy will take the 3-D partition as the first round partition, partition the results again by using a 2-D partition method. This hybrid partition method makes the proposed partition strategy achieve load balance and low communication computation ratio for different algorithms and architectures.

The data placement is closely related to the representation of the graph, and it has a serious effect on system performance.

The upper level framework requires the graph representation method to provide a high memory bandwidth utilization and ensure the locality of memory access as much as possible. While the lower level storage requires high space utilization, avoid the space-wasting for sparsity graph. The storage level also requires the graph that can be loaded into the memory during the I/O operation in graph processing. In order to meet both the upper level and lower level requirements, this article provides a hybrid CSR/CSC graph representation. Then, we further mixed the edge-list representation into the hybrid representation according to the characteristics of the graph. The mixed CSR/CSC graph representation is a benefit for the Scatter/Gather operation. For example, some implementation of the BFS algorithm will change the traversal direction from bottom-up to top-down (top-down to bottom-up), and this hybrid representation will improve the memory access efficiency in this kind of operation. Community is another essential characteristic of the real world graphs, i.e. the vertices in a community are connected but few vertices connect with the vertices outside the community. In this kind of graph, the community can be processed by some SIMD devices, such as GPU, by using the edge-list representation will achieve an excellent performance. Hence, edge-list representation can be an optional method for users.

V. RELATED WORK

We introduce state-of-the-art works for graph computing in this section. Most recent works can be classified into the storage model, the programming model, and the execution model three aspects. We will introduce the related works from these three aspects, receptively.

A. Storage Model

Since most graph applications are memory intensive with a random memory access model, on the other hand, the real-world graphs are with huge size. Both these characteristics will lead to high overhead for both memory and disk access. In order to solve this problem, many researchers proposed a set of state-of-the-art optimization methods. For example, some researchers proposed to use a new storage device, such as Flash-based SSD, to reduce data access overhead. There are also some other optimization methods. In GraphChi [33], the authors proposed a sliding window method to load the graph blocks into memory on demand; by using this method, GraphChi can significantly reduce the disk access overhead. GridGraph [32] proposed a 2-D edge partition method to selectively schedule the graph blocks to reduce the I/O overhead, and the experimental results show the proposed method can achieve a useful data accessing performance. In EC-VHP [34], the authors designed both a simple hash index structure and a multiqueue parallel sequential index structure to improve the processing efficiency of message communication. GraphX [35], which is the core component of Spark [36], providing a stack data solution on top of Spark, can conveniently and efficiently complete a complete set of pipeline operations for graph calculation. Chaos [37] used the secondary memory and graph partitioning scheme to maximize the degree of parallelism and reduce communication overhead.

B. Programming Model

Most of existed research, such Pregel [38], which is the first graph processing system developed by Google, as well as PowerGraph [31], GraphLab [39], and PowerLyra [40], are adopt the vertex-centric. The vertex-centric model-based GPFs are using the Think Like A Vertex (TLAV) paradigm [41]. Compared with the traditional MapReduce model, the TLAV provides a better locality of data access and better scalability, which makes it is more fixable to implement the graph algorithms. However, the overhead of a large amount of random memory access to the TLAV frameworks limits the system performance. In order to solve this problem, X-Stream [42] proposed an edge-centric programming method. In the edge-centric programming model, the edges can be visited in a sequential fashion, which can significantly improve the data access efficiency. In addition, there are some other kinds of the programming model, such as the path-centric programming model proposed in PathGraph [43] and the data-centric programming method [19].

C. Execution Model

Many recent research, such as Gemini [44], Cyclops [45], and Hama [46], are adopt the bulk synchronous parallel (BSP) [38] execution model. In the BSP execution model, a global synchronization is required at the end of each iteration. Due to the uneven degree distribution of the real graph, it is easy to lead to the straggler problem for the vertex-centric programming model because the small degree vertices need to wait for the large degree vertices in the synchronization operation. In order to solve this problem, some other recent state-of-the-art works, such as Chaos [37], PowerGraph [31], PowerLyra [40], and GPS [47], adopt the GAS [31] execution model. In the GAS model, all the operations have been divided into three phases: the information collection phase (Gather), the application phase (Apply), and the distribution phase (Scatter). Since the GAS is an asynchronous execution model, the read-write and write-write conflicts should be considered. There are also some other types of execution models, such as the similar variants of GAS, and GunRock [19] divided the execution operation into Advance, Computer, and Filter phases, while SC-BSP divided the operation as Update, Push, Pull, and Sink (UPPS) [48].

D. Graph Processing on New Architectures

In recent years, there are also some researchers proposed some works, which focused on GPU, FPGA, and some other new architectures. For example, the Medusa [30], GunRock [19], CuSha [22], and Frog [49] are designed on GPU. These works attempt to use the high memory bandwidth to improve the memory access efficiency of GPU, and they also proposed some insights on the programming model, execution model, and data storage model. FPGP [50] and Graphops [51] are designed by using the FPGA. There are also some frameworks designed for the hybrid architectures, such as the CPU and GPU hybrid GPF, i.e., Totem [21]. There are also some attempts to design the hardware-based accelerator for graph processing, such as Graphicionado [52]

that is designed by combating the application execution inefficiencies on general-purpose CPUs, while article [53] proposed a hardware-driven solution.

Due to the booming applications, graph processing has attracted lots of attention from both academia and industry. Though there are lots of state-of-the-art works focused on transitional architecture, it is hard to unleash the computing efficiency of the hardware. There are few works focused on the new architecture devices, and most of the existed works are experimental works, which is hard to program, with low availability. This article attempts to propose an ideal framework for the new architecture devices, which can provide a reference for future works.

VI. CONCLUSION AND FUTURE OPPORTUNITIES

Intelligent data processing for Smart CPSs has attracted much attention from both industry and academics because of the complex dynamic interaction with their environment without any prior information. Focused on the large-scale HITL data processing challenges, this article designed a set of experiments to illustrate the performance of existed GPFs and then proposed a graph-based HITL CPS data processing framework, named Barge, which can fit for different computing devices. By using the semantic model, Barge can implement and map the same code to different computing devices without any change, which can release the programmer from the strenuous and repetitive work. Furthermore, the architecture-driven programming model can make programming logic suitable for kinds of parallel devices, such as GPU, FPGA, ASIC, and many other smart devices. In addition, the data- and topology-aware graph data partition scheme of Barge will partition the large-scale graphs by considering the data structure and also the feature of the computing device to make sure that the large-scale graph to be processed efficiently on smart devices. In further, we will design and implement a hardware compatible framework that can be mapped on to kinds of parallel devices, such as GPU, FPGA, and ASIC, without change the codes.





REFERENCES

- [1] *Fog Computing and the Internet of Things: Extend the Cloud to Where the Things Are*, Cisco, San Jose, CA, USA, 2015.
- [2] M. Saadi, M. TalhaNoor, A. Imran, W. TariqToor, S. Mumtaz, and L. Wuttisittikulij, "Iot enabled quality of experience measurement for next generation networks in smart cities," *Sustain. Cities Soc.*, vol. 60, no. 9, Sep. 2020, Art. no. 102266.
- [3] 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.
- [4] M. H. Cintuglu, O. A. Mohammed, K. Akkaya, and A. S. Ulugac, "A survey on smart grid cyber-physical system testbeds," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 446–464, 1st Quart., 2017.
- [5] J. Sztipanovits *et al.*, "Toward a science of cyber-physical system integration," *Proc. IEEE*, vol. 100, no. 1, pp. 29–44, Jan. 2012.
- [6] F. Hofer, "Architecture, technologies and challenges for cyber-physical systems in industry 4.0: A systematic mapping study," in *Proc. 12th ACM/IEEE Int. Symp. Empirical Softw. Eng. Meas.*, New York, NY, USA, 2018, pp. 1–10.
- [7] M. I. Ashraf, M. Bennis, C. Perfecto, and W. Saad, "Dynamic proximity-aware resource allocation in vehicle-to-vehicle (V2V) communications," in *Proc. IEEE Globecom Workshops*, Dec. 2016, pp. 1–6.
- [8] S. Mumtaz, A. Gameraio, and J. Rodriguez, "EESM for IEEE 802.16e: WiMaX," in *Proc. 7th IEEE/ACIS Int. Conf. Comput. Inf. Sci.*, May 2008, pp. 361–366.
- [9] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, and L. T. Yang, "Data mining for Internet of Things: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 77–97, Qua. 2014.
- [10] Y. Liu, X. Fang, M. Xiao, and S. Mumtaz, "Decentralized beam pair selection in multi-beam millimeter-wave networks," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2722–2737, Jun. 2018.
- [11] J. Wu, S. Guo, J. Li, and D. Zeng, "Big data meet green challenges: Big data toward green applications," *IEEE Systems J.*, vol. 10, no. 3, pp. 888–900, Sep. 2016.
- [12] J. Wu, S. Guo, H. Huang, W. Liu, and Y. Xiang, "Information and communications technologies for sustainable development goals: State-of-the-art, needs and perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2389–2406, 3rd Quart., 2018.
- [13] K. Zhang, L. Lan, J. T. Kwok, S. Vucetic, and B. Parvin, "Scaling up graph-based semisupervised learning via prototype vector machines," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 3, pp. 444–457, Mar. 2015.
- [14] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proc. IEEE*, vol. 104, no. 1, pp. 11–33, Jan. 2016.
- [15] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, doi: 10.1109/TKDE.2020.2981333.
- [16] Y. Lou, M. Uddin, N. Brown, and M. Cafarella, "Knowledge graph programming with a human-in-the-loop: Preliminary results," in *Proc. Workshop Hum. Loop Data Anal.*, New York, NY, USA, 2019, pp. 1–7.
- [17] A. Holzinger *et al.*, "Interactive machine learning: Experimental evidence for the human in the algorithmic loop," *Int. J. Speech Technol.*, vol. 49, no. 7, pp. 2401–2414, Jul. 2019.
- [18] A. Doan, "Human-in-the-loop data analysis: A personal perspective," in *Proc. Workshop Human Loop Data Analy.*, New York, NY, USA, 2018, pp. 1–5.
- [19] Y. Wang, A. Davidson, Y. Pan, Y. Wu, A. Riffel, and J. D. Owens, "Gunrock: A high-performance graph processing library on the GPU," in *Proc. 20th ACM SIGPLAN Symp. Princ. Pract. Parallel Program.*, New York, NY, USA, 2015, pp. 1–12.
- [20] H. Wang, L. Geng, R. Lee, K. Hou, Y. Zhang, and X. Zhang, "SEP-graph: Finding shortest execution paths for graph processing under a hybrid framework on GPU," in *Proc. 24th Symp. Princ. Pract. Parallel Program.*, New York, NY, USA, Feb. 2019, p. 38.
- [21] A. Gharaibeh, L. Beltr ao Costa, E. Santos-Neto, and M. Ripeanu, "A yoke of oxen and a thousand chickens for heavy lifting graph processing," in *Proc. 21st Int. Conf. Parallel architectures compilation Techn.*, New York, NY, USA, 2012, pp. 345–354.
- [22] F. Khorasani, K. Vora, R. Gupta, and L. N. Bhuyan, "CuSha: Vertex-centric graph processing on GPUs," in *Proc. 23rd Int. Symp. High-Perform. Parallel Distrib. Comput.*, New York, NY, USA, 2014, p. 239–252, p. 239.
- [23] A. H. Nodehi Sabet, J. Qiu, and Z. Zhao, "Tigr: Transforming irregular graphs for GPU-friendly graph processing," in *Proc. Int. Conf. Archit. Support Program. Lang. Oper. Syst.*, New York, NY, USA, 2018, p. 622–636.
- [24] K. Meng, J. Li, G. Tan, and N. Sun, "A pattern based algorithmic autotuner for graph processing on GPUs," in *Proc. 24th Symp. Princ. Pract. Parallel Program.*, New York, NY, USA, Feb. 2019, p. 201–213.
- [25] D. B. West, *Introduction to Graph Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.
- [26] J. Leskovec. (2009). *Stanford Network Analysis Project*. <http://snap.stanford.edu/index.html>
- [27] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Nov. 2009.
- [28] Y. Virkar and A. Clauset, "Power-law distributions in Binned empirical data," *Ann. Appl. Statist.*, vol. 8, no. 1, pp. 89–119, Mar. 2014.
- [29] S. Kanev *et al.*, "Profiling a warehouse-scale computer," *SIGARCH Comput. Archit. News*, vol. 43, no. 3S, p. 158–169, Jun. 2015.
- [30] J. Zhong and B. He, "Medusa: Simplified graph processing on GPUs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 6, pp. 1543–1552, Jun. 2014.
- [31] J. E. Gonzalez, Y. Low, H. Gu, D. Bickson, and C. Guestrin, "Powergraph: Distributed graph-parallel computation on natural graphs," in *Proc. 10th USENIX Conf. Operating Syst. Design Implement.*, 2012, p. 17–30.
- [32] X. Zhu, W. Han, and W. Chen, "Gridgraph: Large-scale graph processing on a single machine using 2-level hierarchical partitioning," in *Proc. 2015 USENIX Conf. Usenix Annu. Tech. Conf.*, 2015, p. 375–386.

- [33] A. Kyrola, G. Blelloch, and C. Guestrin, "Graphchi: Large-scale graph computation on just a PC," in *Proc. 10th USENIX Conf. Oper. Syst. Design Implement.*, 2012, p. 31–46.
- [34] L. Fangling *et al.*, "Edge cluster based large graph partitioning and iterative processing in BSP," *J. Comput. Res. Develop.*, vol. 52, no. 4, p. 960, 2015.
- [35] J. E. Gonzalez, R. S. Xin, A. Dave, D. Crankshaw, M. J. Franklin, and I. Stoica, "Graphx: Graph processing in a distributed dataflow framework," in *Proc. 11th USENIX Conf. Oper. Syst. Des. Implement.*, 2014, p. 599–613.
- [36] M. Zaharia *et al.*, "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing," in *Proc. 9th USENIX Conf. Netw. Syst. Design Implement.*, 2012, p. 2.
- [37] A. Roy, L. Bindschaedler, J. Malicevic, and W. Zwaenepoel, "Chaos: Scale-out graph processing from secondary storage," in *Proc. 25th Symp. Operating Syst. Princ.*, New York, NY, USA, 2015, pp. 410–412.
- [38] G. Malewicz *et al.*, "Pregel: A system for large-scale graph processing," in *Proc. Int. Conf. Manage. Data*, New York, NY, USA, 2010, p. 6.
- [39] Y. Low, J. Gonzalez, A. Kyrola, D. Bickson, C. Guestrin, and J. Hellerstein, "Graphlab: A new framework for parallel machine learning," in *Proc. Conf. Uncertainty Artif. Intell.*, Arlington, VA, USA, 2010, p. 340–349.
- [40] R. Chen, J. Shi, Y. Chen, B. Zang, H. Guan, and H. Chen, "Powerlyra: Differentiated graph computation and partitioning on skewed graphs," *ACM Trans. Parallel Comput.*, vol. 5, no. 3, p. 15, Jan. 2019.
- [41] R. R. McCune, T. Weninger, and G. Madey, "Thinking like a vertex: A survey of vertex-centric frameworks for large-scale distributed graph processing," *ACM Comput. Surv.*, vol. 48, no. 2, p. 12, Oct. 2015.
- [42] A. Roy, I. Mihailovic, and W. Zwaenepoel, "X-stream: Edge-centric graph processing using streaming partitions," in *Proc. ACM Symp. Operating Syst. Princ.*, New York, NY, USA, 2013, pp. 472–488.
- [43] P. Yuan, W. Zhang, C. Xie, H. Jin, L. Liu, and K. Lee, "Fast iterative graph computation: A path centric approach," in *Proc. Int. Conf. for High Perform. Comput., Netw., Storage Anal.*, Nov. 2014, p. 401.
- [44] X. Zhu, W. Chen, W. Zheng, and X. Ma, "Gemini: A computation-centric distributed graph processing system," in *Proc. 12th USENIX Conf. Operating Syst. Des. Implement.*, 2016, p. 301–316.
- [45] R. Chen, X. Ding, P. Wang, H. Chen, B. Zang, and H. Guan, "Computation and communication efficient graph processing with distributed immutable view," in *Proc. 23rd Int. Symp. High-Perform. Parallel Distrib. Comput.*, New York, NY, USA, 2014, pp. 215–226.
- [46] S. Seo, E. J. Yoon, J. Kim, S. Jin, J.-S. Kim, and S. Maeng, "HAMA: An efficient matrix computation with the MapReduce framework," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Technol. Sci.*, Nov. 2010, pp. 721–726.
- [47] S. Salihoglu and J. Widom, "GPS: A graph processing system," in *Proc. 25th Int. Conf. Scientific Stat. Database Manage.*, New York, NY, USA, 2013, pp. 1–4.
- [48] Z. Xiang, L. Bo, S. Haichuan, and X. Weidong, "A revised BSP-based massive graph computation model," *Chin. J. Comput.*, vol. 40, no. 1, pp. 223–235, Jan. 2017.
- [49] X. Shi *et al.*, "Frog: Asynchronous graph processing on GPU with hybrid coloring model," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 1, pp. 29–42, Jan. 2018.
- [50] G. Dai, Y. Chi, Y. Wang, and H. Yang, "FPGP: Graph processing framework on FPGA a case study of breadth-first search," in *Proc. ACM/SIGDA Int. Symp. Field-Program. Gate Arrays*, New York, NY, USA, 2016, p. 10–105.
- [51] T. Oguntebi and K. Olukotun, "GraphOps: A dataflow library for graph analytics acceleration," in *Proc. ACM/SIGDA Int. Symp. Field-Program. Gate Arrays - FPGA*, New York, NY, USA, 2016, p. 111–117.2016, p. 111.
- [52] T. J. Ham, L. Wu, N. Sundaram, N. Satish, and M. Martonosi, "Graphicionado: A high-performance and energy-efficient accelerator for graph analytics," in *Proc. 49th Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO)*, Oct. 2016, pp. 1–4.
- [53] M. M. Ozdal *et al.*, "Energy efficient architecture for graph analytics accelerators," in *Proc. ACM/IEEE 43rd Annu. Int. Symp. Comput. Archit. (ISCA)*, Jun. 2016, p. 166–177.



An AI-enabled lightweight data fusion and load optimization approach for Internet of Things

Mian Ahmad Jan^a  , Muhammad Zakarya^a  , Muhammad Khan^b, Spyridon Mastorakis^c, Varun G. Menon^d, Venki Balasubramanian^e, Ateeq Ur Rehman^a

Show more 

 Outline |  Share  Cite

<https://doi.org/10.1016/j.future.2021.03.020> 

[Get rights and content](#) 

Highlights

- A lightweight data fusion approach using stratified sampling.
- A dynamic load optimization approach using Evolutionary algorithms to maintain balanced traffic.
- A dynamic service migration technique to balance the load across several edge servers that triggers migration decisions.

Abstract

In the densely populated Internet of Things (IoT) applications, sensing range of the nodes might overlap frequently. In these applications, the nodes gather highly correlated and redundant data in their vicinity. Processing these data depletes the energy of nodes and their upstream transmission towards remote datacentres, in the fog infrastructure, may result in an unbalanced load at the network gateways and edge servers. Due to heterogeneity of edge servers, few of them might be overwhelmed while others may remain less-utilized. As a result, time-critical and delay-sensitive applications may experience excessive delays, packet loss, and degradation in their Quality of Service (QoS). To ensure QoS of IoT applications, in this paper, we eliminate correlation in the gathered data via a lightweight data fusion approach. The buffer of each node is partitioned into strata that broadcast only non-correlated data to edge servers via the network gateways. Furthermore, we propose a dynamic service migration technique to reconfigure the load across various edge servers. We assume this as an optimization problem and use two meta-heuristic algorithms, along with a migration approach, to maintain an optimal Gateway-Edge configuration in the network. These algorithms monitor the load at each server, and once it surpasses a threshold value (which is dynamically computed with a simple machine learning method), an exhaustive search is performed for an optimal and balanced periodic reconfiguration. The experimental results of our approach justify its efficiency for large-scale and densely populated IoT applications.

PDF

Help



Keywords

Internet of Things; Data fusion; Load optimization; Evolutionary algorithms; Gateway-Edge configuration; Service migration

1. Introduction

In the Internet of Things (IoT), the sensor nodes of various applications gather highly correlated data in their neighbourhoods that affect the outcome of any decision made at the cloud data centres[1], [2]. In these applications, the data are unstructured, intermittent and somewhat dynamic. The raw data gathered by the nodes need to be processed locally and analysed at the edge and cloud data centres to optimize the usage of available resources. The raw data need to be fused within the network to reduce the correlation in them. Each node, unaware of its neighbour's sensing range, gathers data in its neighbourhood. The sensing range of two or more nodes may overlap leading to the aggregation of similar data[3]. Each node needs to perform local data fusion to discard multiple copies of the same data. In-network data fusion alleviates the redundancy to trade-off the volumes of data and the available resources at the edge and cloud data centres[4]. The presence of resource-starving nodes means that a data fusion approach needs to be lightweight, robust, and scalable, based on application requirements.

Data fusion alone is not enough to optimize the usage of available network resources. The upstream fused data toward the cloud data centres need to be fairly distributed among the edge servers[5], [6]. In the existing literature[7], [8], [9], the fused data streams are offloaded to the nearest edge servers. However, this approach is not efficient as some of these servers may overload quickly in comparison to others that remain underutilized. The underlying nodes and network gateways associated with the over-utilized servers may suffer higher latency, packet drop, and bandwidth consumption. For a fair distribution of the network load, a dynamic load balancing approach needs to be adopted to assign the time-consuming tasks to underutilized servers. Based on the run-time load at the servers, a decision needs to be made for the assignment of data streams. The configuration of network entities needs to be constantly monitored for an optimized and balanced load. Artificial Intelligence (AI)-enabled algorithms can manage the complex relationship among the network entities[10], [11]. These algorithms need to be adopted for intelligent load balancing and optimization of the selected paths for reliable transmission of the fused data. They have the ability to reconfigure the devices' connectivity based on their experienced load.

Moreover, heterogeneity of the edge servers and network bandwidth may generate opportunities for application migrations (running within virtual machines) which could be beneficial in further load-balancing, avoiding stranded (wasted) resources, and performance degradation (due to overload situations). Stranded resources are those which cannot be allocated due to the unavailability of another resource e.g. CPU cores are fully utilized but memory is half utilized – half memory cannot be allocated because there are no CPU cores available to run the VM/workload/application. Here, heterogeneity refers to the speed of server to process data or network bandwidth. This is achieved through comparing the current utilization levels of the edge servers and/or the rate of transferring over a network link (channel conditions) to some pre-defined threshold values. If utilization level of an edge server or a network link surpasses a particular threshold value, migrations will happen. However, a static threshold may not be appropriate; therefore, we use a simple machine learning model to compute an adaptive threshold.

In this paper, we propose a novel data fusion and load optimization approach for the IoT-enabled applications. Our approach reduces data redundancy at the node level and fairly distributes the fused data streams among the edge servers. It is scalable and can be used by any application, provided that the threshold values for monitored data are known. It ensures the availability of high-quality data at the cloud data centres for decision-making. The main contributions of this work are as follows.

1. A lightweight data fusion approach that reduces the correlation and redundancy in the gathered data by using *MiniMax* stratified sampling. The buffer of each node is partitioned into multiple stratum, each one holding only two values, i.e., a minimum (*min*) and a maximum (*max*). A comparison with *min* and *max* decides to discard or retain any newly sensed data. After a sampling interval, the stratum of each node transmits only two data readings by discarding all other correlated readings.
2. A dynamic load optimization approach that maintains a balanced traffic in the network using a real-valued Genetic Algorithm (GA) and Discrete Particle Swarm Optimization (DPSO). A Software-Defined Networking (SDN) controller monitors the load on individual edge servers and reconfigures the current Gateway-Edge configuration if an unbalanced load is experienced. For reconfiguration, the SDN controller invokes these evolutionary algorithms to identify the transmission path for each gateway towards a prospective server.
3. The above contribution does not account for dynamic load-balancing, i.e., when on some particular resources, the data get processed quicker than others. A dynamic service migration technique is suggested to balance the load across several edge servers that triggers migration decisions, based on current resource (edge server, network channel) usage. A dynamic threshold is computed using a simple regression model in order to keep resources well-balanced.

The rest of our paper is organized as follows. In Section 2, we provide an overview of the background studies pertaining to our proposed approach. In Section 3, our proposed framework and algorithms are described in detail. This section also offers a service migration technique for load balancing across several edges. The experimental results and performance evaluation are sketched in Section 4. Finally, we provide concluding remarks and future research directions in Section 5.

2. Background

In this section, we provide the background studies pertaining to data fusion in the context of load optimization for IoT applications.

In [12], a cloud-based adaptive sensing belief propagation protocol (ASBP) was proposed. ASBP estimates the quality of links to determine the shortest routes toward the cloud for data gathered from IoT applications. The protocol exploits the spatio-temporal correlation among the data streams at cloud datacentres to reduce the energy consumption, and balance the load by keeping a subset of nodes in active states at a given time. ASBP, however, is unable to evenly distribute the load on edge servers for a large-scale IoT network. Besides, fusing massive amount of sensor data at the cloud incur a significant amount of transmission overhead. A dynamic sensor activation algorithm, SensorRank, was proposed to prioritize the deployed nodes based on their residual energy levels, their relative distance, and their links qualities [13]. SensorRank considered symmetric channels for data transmission among the neighbouring nodes. These channels may lead to an uneven load distribution among the nodes, and on the gateways and edge servers, respectively. A spatio-temporal based novel data mining approach (NDM) was proposed for the removal of redundant data, prior to upstream transmission towards the gateways [14]. NDM uses a packet classification approach to filter out redundant data to maintain the network load on the edge servers. NDM is non-scalable and its iterative nature of load distribution at the edge incurs an excessive overhead at the resource constrained sensor nodes.

In [11], an optimized mobile sink-based load balancing (OMS-LB) protocol was proposed to achieve balanced load for a large-scale IoT network. OMS-LB offloads the computationally complex tasks from data gathering devices to a Software-defined Network (SDN) controller that is interfaced with cloud datacentres. The proposed protocol uses PSO and GA to determine the optimal paths for a mobile edge server and optimal data gathering points, i.e., gateways. OMS-LB does not define any criteria for data collection from an application perspective. Besides, the presence of a single server makes this protocol non-scalable, and vulnerable to security threats. A multi-edge based architecture was proposed for seamless integration of cloud datacentres in an IoT environment [5]. The proposed architecture used a multilevel protocol for gateways selection and AI-based load balancer for the identification of an optimal load distribution. However, the proposed architecture lacks any information about the heterogeneity of

nodes, network latency and bandwidth requirements. In [15], the authors proposed a data aggregation scheme by estimating an accurate sensor matrix from the gathered raw data. A fog server is used to reconstruct the matrix that contains minimal noise and highly refined data. However, the proposed matrix does not take into account the load balancing issue and has limitations imposed on its scalability. Besides, it lacks any information on heterogeneous data fusion and interoperability of IoT devices.

All these existing approaches focused on centralized gateways and edge servers for load optimization and decision-making. The presence of centralized entities affect the scalability, fault tolerance and optimal load adjustment of a network. Besides, these approaches operate without data aggregation and fusion at the network level. As a result, they require excessive processing and storage of redundant data at the network gateways and edge servers. The transmission of redundant data ultimately deteriorate the QoS of an underlying network. In the IoT paradigms, it is inevitable to consider AI algorithms for maintaining a balanced load for various applications. There are numerous AI algorithms developed to resolve the load optimization problem. However, in this paper, we utilize the most embraced evolutionary algorithms, i.e., Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) [16], [17], [18]. GA and PSO are population-based algorithms, where the population means a group of all possible solutions, i.e., Gateway-Edge configurations (load balancing and optimization) [19].

GA is a bio-inspired search algorithm in which the population is referred to as a group of chromosomes. The genes of the fittest available chromosomes are utilized to generate new chromosomes, i.e., new optimal Gateway-Edge configurations, via mutation and crossover. On the other hand, in standard PSO, the population of all possible solutions is referred to as a swarm of particles. PSO is inspired by the social behaviour of swarms of ants, a flock of birds, a shoal of fish, etc. In all these cases, the swarm probe the search space for identifying the food with varying velocities. In the case of PSO, each particle is considered a candidate solution for the Gateway-Edge configuration problem. In the case of GA, each chromosome is considered a candidate solution. Since both these algorithms are not directly applicable to integer-based load optimization problems, we have developed a real-valued GA and a Discrete PSO (DPSO) for identifying the optimal Gateway-Edge configurations.

Load balancing is an essential part of the IoT, edge and cloud frameworks that could be achieved in two different ways: (i) dynamic service placement; and (ii) service migration. In respect of (i), two policies are suggested in [20]: cloud-only placement: place all application's modules in the server; and edge-ward placement: favour to run application's modules on various edge devices. Moreover, if allocation of an edge device is not suitable for a particular module, then either resources from other edge devices (server) could be provisioned or it could be migrated somewhere else. Empirical evaluation of both policies suggests that the edge-ward policy significantly improves the application's performance and reduces the network traffic. In respect of (ii), authors in [21], [22] suggest that if an application's performance is the worst on a particular edge device (due to more number of connected sensor devices, network congestion etc.), then its migration either to the server or to another edge device could improve its performance and reduces network traffic. Moreover, mobility management in mobile edge clouds (MECs) also involves migrations [22].

Migrations could also be triggered to balance resource utilization levels of edge nodes. For example, if the utilization of an edge node increases certain threshold value (say 80%), some of the application's module may be moved to other edges. Migration can also take place when resources are under utilized i.e. threshold of 20%. This is done to conserve and consolidate resources to save energy [23]. In the later case, energy could be saved through migrating workloads from these underutilized servers to other servers; and switching them off. However, this may cause performance issues, in particular, if demand exceeds suddenly. We, in this paper, prefer the former one as our objective is not saving energy; instead we want to balance the load across different switched on servers. Furthermore, we use a dynamic threshold-based method that estimate these threshold values periodically – using Eq. (12). Service migration could only be achieved if various sand-boxing technologies such as virtualization, containerization are being used to virtualize the server and edge device resources [24]. In practice, resources in public clouds are virtualized, which increases resource utilization levels and saves energy. If various modules of a particular application are being run in a Virtual Machine (VM) or container; then the service can be migrated either off-line or live. In live migration, the service is moved transparently while still running; however, in off-line

migration the service is stopped first, moved, and then resumed at the target edge. Using CRIU¹ technology, containers could be more quickly migrated than VMs. In case of live VM migration, where VMs data are kept on a shared storage reachable over the network, the time of migration T_{mig} depends on the total volume of memory used by the VM M_{vm} and available network bandwidth B_{total} . For virtual machine V_i the total migration time is given by:

$$T_{mig_{V_i}} = \frac{M_{V_i}}{B_{total}} \quad (1)$$

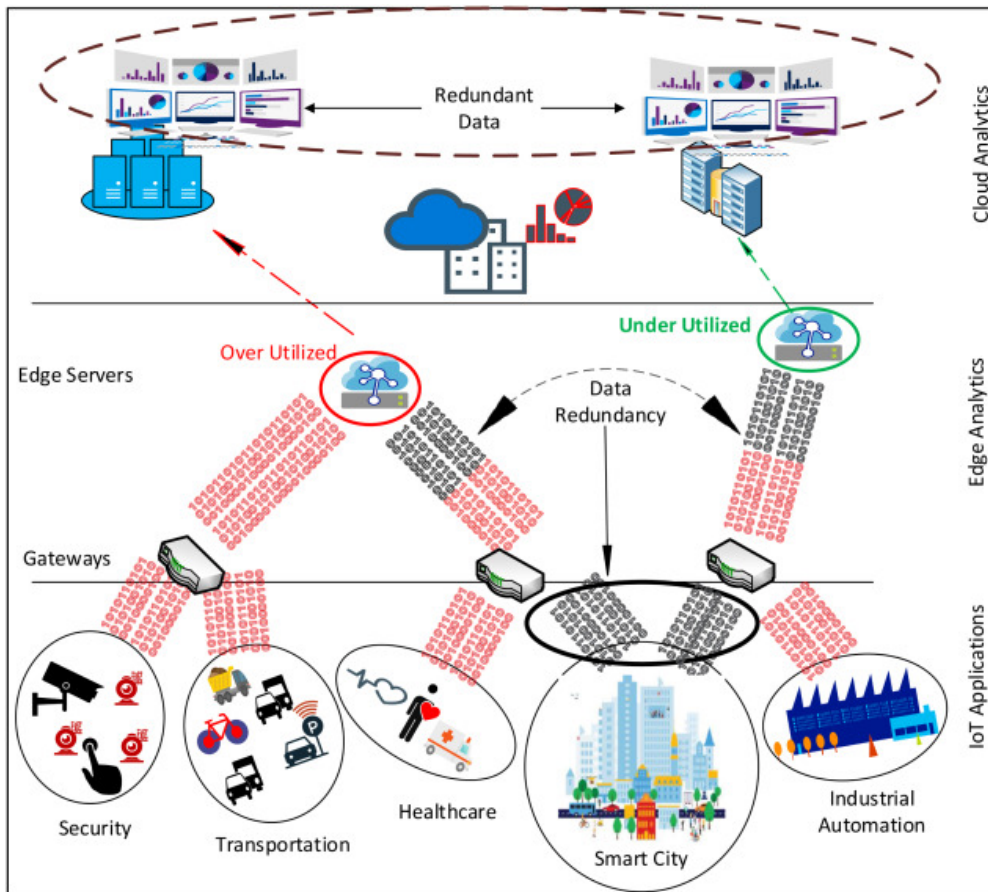
The above equation is used to compute only the migration time of a particular VM. Every VM for this T_{mig} time is considered offline, which is also called the downtime of the VM. The downtime (or performance loss) is dependent on the migration duration, as given by Eq.(1)[25]. Increased downtime results in poor performance; therefore it should be minimized for high availability of the datacenter. The performance degradation P_{deg} due to a single migration is calculated using the following formula (as given by Eq.(2)), where U_{V_i} is the CPU utilization of VM V_i , t_0 is the migration start time and 0.1 is the factor that shows the average performance degradation for web application i.e. 10% of the CPU utilization[23], [26].

$$P_{deg_{V_i}} = 0.1 \times \int_{t_0}^{t_0+T_{mig_{V_i}}} U_{V_i}(t) \cdot dt \quad (2)$$

Note that, the above performance degradation model (10% loss in workload execution time) is benchmarked in[23], [27]; and we assume that it already accounts for other time consuming activities such as: the time to initiate a VM migration; the time to transfer page files (dirty pages in case of live or online VM migration); the time to boot/spin up a new server (if there is no currently running server that can accept the VM being migrated); and the time to restart the VM (in the case of cold or offline VM migration)[28].

3. Data fusion and load optimization approach for IoT applications

In Fig.1, the sensor nodes of various applications transmit their data to cloud data centres via the network gateways and edge servers. Among these applications, the smart city nodes gather and transmit highly correlated data streams. The transmission of these streams affects the decision-making at data centres and creates bandwidth bottlenecks for time-critical and delay-sensitive applications. Moreover, these applications experience excessive latency and degradation in the network throughput if an unbalanced load is experienced at the edge servers. An uneven load distribution results in some of the servers over-utilized while others remain underutilized. The unbalanced load leads to packet loss, longer delays, and network congestion. In this section, we discuss our proposed data fusion and load optimization approach to eliminate data redundancy and maintain a balanced load at the network entities.



[Download : Download high-res image \(1MB\)](#)

[Download : Download full-size image](#)

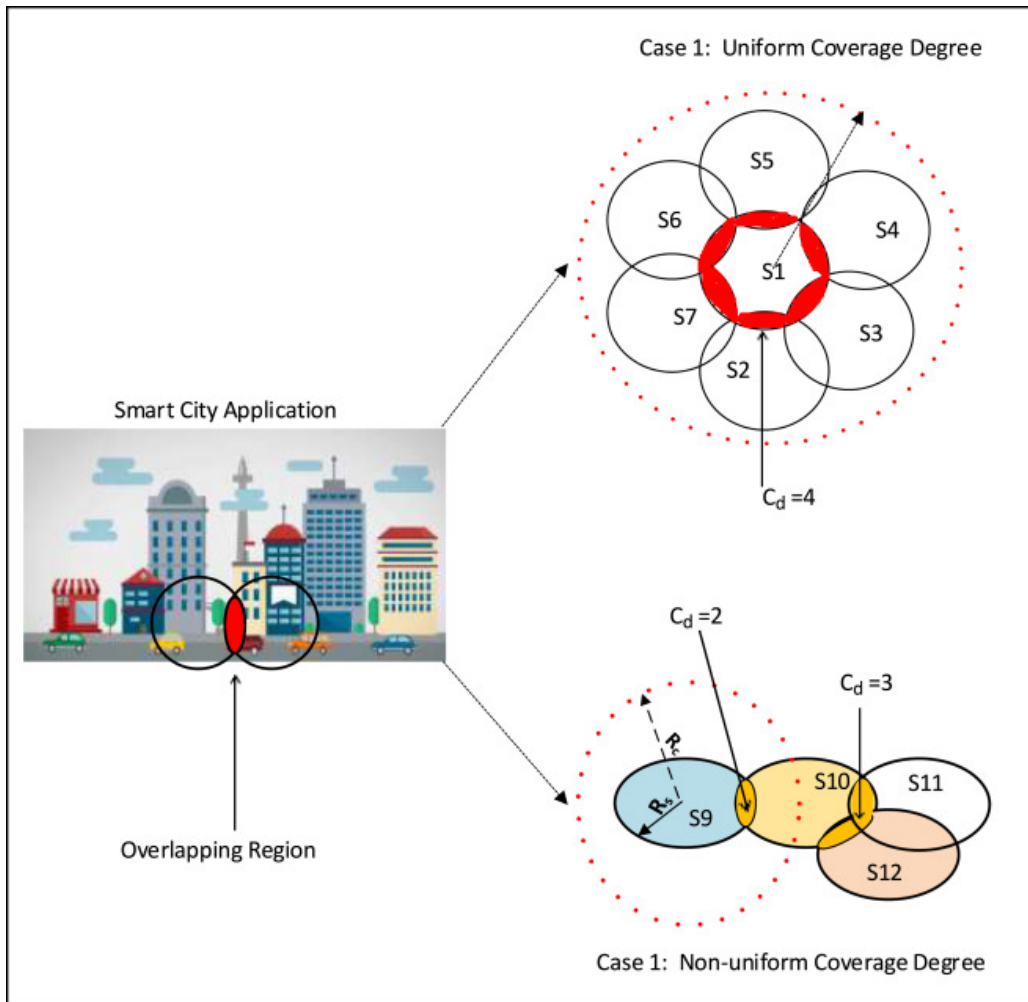
Fig. 1. Data redundancy and unbalanced load in IoT.

3.1. The Proposed Fog-IoT Framework

The proposed model consists of three layers i.e. the cloud layer, the edge layer and the local layer, as shown in Fig. 1. The local layer is responsible to gather important data (related to traffic, healthcare, and crowd, etc.) through various IoT devices and sensors. Once the data is collected, it is processed and/or stored at the edge layer through edge clouds [29]. The edge level processing may also include aggregation that could be achieved through removing redundant data. The filtered data can, then, be sent to the remote cloud layer for further processing such as storage, monitoring and resource management. Transferring the gathered data at local layer directly to the cloud layer, or through edge, may introduce significant delays in the cloud network, which is optimized through data fusion and load balancing methods as discussed in Sections 3.2 Data fusion, 3.3 Load optimization.

The edge infrastructure is of great use when reading the stored data for processing through machine learning approaches. For example, real-time prediction of the traffic flow might happens at the edge layer, however, prediction for monitoring services (load, service migration) can be performed at the cloud layer [30]. Moreover, if real-time prediction, for example, shortest or safe route estimation, is carried out on the remote cloud, then it will incur significant delays depending on the network quality and capacity. In that scenario, the nearest edge cloud can predict the road conditions, congestion and distance; if the required data is stored locally. However, due to the least storage and processing capabilities of the edge clouds [31], the data may not be available or processed locally. In that case, there are three various options: (i) move the required data from the cloud to the edge, process, take decisions, and discard; (ii) perform the prediction at the remote cloud (in whole); and (iii) train the prediction model at the remote cloud and predict at edge layer (distributed fashion computation). Similarly, the huge amount of collected data may consist of duplicate values that could create network congestion and, therefore, affect the prediction process. The edge cloud can use fusion and aggregation technique to send only appropriate data to the remote cloud.

PDF
Help



[Download : Download high-res image \(591KB\)](#)

[Download : Download full-size image](#)

Fig. 2. Data Correlation of varying coverage degrees.

3.2. Data fusion

In a densely deployed smart city, energy hole problem is a common issue faced by the one-hop neighbours of the base station [32]. These one-hop neighbours not only transmit their own data but also relay the data of downstream nodes to the base station. As a result, their energy is depleted rapidly as compared to other nodes. To fill the void left by energy-depleted nodes and to maintain seamless network connectivity, one or more nodes may either sense multiple regions or move around the field to fill this gap. These nodes continuously sense and aggregate data in their neighbourhood, as shown in Fig.2. Each node S maintains a coverage area based on its sensing range (R_s), and a radio coverage based on its communication range (R_c), respectively. The R_s enables efficient data monitoring, whereas the R_c ensures the upstream data transmission. These nodes can have a uniform or non-uniform coverage degree. The coverage degree represents the number of nodes actively monitoring a particular region, i.e., an overlapped region. For uniform coverage, the value of correlation degree (C_d) remains constant for all the nodes. On the other hand, the value of C_d varies for some or all the nodes in a non-uniform coverage. A larger value of C_d represents highly correlated and redundant data as multiple nodes monitor a particular event in the overlapped region.

To eliminate the correlated and redundant data, we use a lightweight data fusion approach at the node level. The proposed approach uses a *minimax* function for the identification and removal of redundant data. In a smart city, each node is equipped with multiple sensors based on an application requirement. In this paper, we restrict our discussion to the temperature sensors only. However, the flexibility of our approach enables it to be extended for any application provided that the threshold values of monitored data are known. We classify the sensed

temperature readings based on the correlation and similarity index among them. Each class, also known as stratum, contains a particular range of temperature readings. For each node, we define ten different stratum that are dynamic and depend on application requirements. They are designed using the concept of stratified sampling, a probability-based sampling technique [33], [34].

The strata² are defined within the buffer of each node and can store temperature readings ranging from 20°C to 39.99°C. Based on these readings, each stratum holds a different range of varying values of up to 2°C. For instance, the range of first stratum S_{t_1} is $\{20.00, \dots, 21.99\}$, and the last stratum $S_{t_{10}}$ is $\{38.00, \dots, 39.99\}$, respectively. The outcome of each stratum can either be a *min* or *max* value. Each stratum has a mean value m that defines its *min* and *max*, respectively. In the beginning, when a new temperature reading T_i is sensed by a node, it is checked against strata of the given node to identify a destination stratum. Once a match is found, T_i is compared against m of the stratum. If T_i is less than m , it becomes the *min*, otherwise, it becomes the *max*, as shown in Eq.(3). The next time a new reading T_{i+1} is sensed within the range of the same stratum, it is compared against m . If T_{i+1} is less than m , a comparison is made with the new *min*. If T_{i+1} is less than *min*, the former turns out to be the new *min*, otherwise, it is discarded. A similar comparison is made with *max*. If T_{i+1} is greater than *max*, it turns out to be the new *max*, otherwise, it is discarded. Irrespective of T_i , T_{i+1} or any other subsequent readings, an exact match with the values of m , *min* or *max* means that these readings will be discarded.

$$f(S_{t_1}, \dots, S_{t_{10}}) = \begin{cases} \min, & \text{if } T_i < m, \\ 0, & \text{if } T_i == m, \\ \max, & \text{if } T_i > m. \end{cases} \quad (3)$$

The *max* and *min* of a given stratum can be plotted as a stationary point on a curve. A point $P(x_0, f(x_0))$ is considered a stationary point of a function $f(x)$ if $(\frac{df}{dx})$ is 0 at $x=x_0$. Suppose a function $y=f(x)$ is a stationary point with $x=x_0$. Then

- if $\left[\frac{d^2 f}{dx^2}\right]_{x=x_0} < 0$, then $x=x_0$ is the *max* of a stratum.
- if $\left[\frac{d^2 f}{dx^2}\right]_{x=x_0} > 0$, then $x=x_0$ is the *min* of a stratum.
- if $\left[\frac{d^2 f}{dx^2}\right]_{x=x_0} == 0$, then
 - if $\left[\frac{df}{dx}\right]_{x=x_0} < 0$ for $x > x_0$, and $\left[\frac{df}{dx}\right]_{x=x_0} > 0$ for $x < x_0$, then $x=x_0$ is the *max* of the given strata.
 - if $\left[\frac{df}{dx}\right]_{x=x_0} > 0$ for $x > x_0$, and $\left[\frac{df}{dx}\right]_{x=x_0} < 0$ for $x < x_0$, then $x=x_0$ is the *min* of the given strata.

In our data fusion approach, the sampling rate of each node is S_r packets per second, where $S_r \geq 1$. The stratum of each node transmits only two packets, i.e., a *min* and a *max*, after every one minute. If a node constantly maintains its sampling rate at $S_r = 1$ for one minute, our approach achieves a maximum of 3 times reduction in the number of transmitted packets to the gateways. For $S_r > 1$, the number of transmitted packets is reduced even further minimum of 3 times. In our approach, the S_r of a node and the number of transmitted packets from its strata to gateways are inversely proportional to each other. These strata significantly reduce data redundancy, network latency, packet collision probability, and ultimately the network congestion. In our proposed approach, each node maintains a similarity index (Ω) for the data gathered over the S_r interval. The value of Ω ranges from 0.03 to 0.1. If Ω is equal to 0.03, it means that among two temperature packets within the range of 0.03°C, only one will be retained. For example, in case of two packets with values 20°C and 20.03°C, only one will be retained in the stratum. Hence, larger the value of Ω , higher will be the rate at which the data are fused.

3.3. Load optimization

Upon data fusion, each node transmits the refined data to cloud data centres via the network gateways and edge servers. The gateways are relay nodes that need to be monitored for maintaining a balanced load at the edge servers. For this purpose, an optimal Gateway-Edge configuration is required. We use various Key Performance

Indicators (KPIs) for an in-depth analysis of the network traffic to identify the optimal configuration. An SDN controller is used for identifying the transmission route for each gateway. It monitors the load on each server, and once it surpasses a threshold value, an alarm is raised to re-configure the current Gateway-Edge connection. If the Gateway-Edge configuration is known at a particular time t , then finding the optimal balanced Gateway-Edge configuration at time $t + 1$ is a primary challenge. If N is the number of gateways, and M is the number of edge servers, then the Gateway-Edge configuration at a particular time t can be represented by a vector $G^t = \{G_1^t, G_2^t, \dots, G_n^t\}$, where $G_n^t \in \{1, 2, \dots, N\}$. As an example, $G_n^t = m$ means that the n th gateway is transmitting to an m th server at time t . Finding the optimal Gateway-Edge configuration vector at time $t + 1$, i.e., $G^{t+1} = \{G_1^{t+1}, G_2^{t+1}, \dots, G_n^{t+1}\}$, is a prime objective. To solve the Gateway-Edge configuration problem, we consider two KPIs, i.e., Average Residual Energy (KPI_{ARE}) of the network and Load Fairness Index (KPI_{LFI}) of the servers.

The LFI is monitored based on Jain's Fairness Index[35]. The normalized weighted sum of these two KPIs is taken into account to maximize the network performance (NP) at time t as shown in Eq.(4).

$$Max(NP) = \alpha KPI_{ARE} + \beta KPI_{LFI}. \quad (4)$$

Here, NP is a primary objective function for optimization problems, and α and β are the weights assigned to each KPI. These weights represent the priority level of each KPI in the objective function, as shown in Eq.(5).

$$NP = \alpha \left(\frac{1}{N_n} \sum_{i=1}^{N_n} \frac{R_i(t)}{\hat{E}} \right) + \beta \left(\frac{1}{M} \frac{\left(\sum_{i=1}^M \sum_{n=1}^N I_{n,i} \varphi_n(t) \right)^2}{\sum_{i=1}^M \left(\sum_{n=1}^N I_{n,i} \varphi_n(t) \right)^2} \right). \quad (5)$$

where, $\frac{R_i(t)}{\hat{E}}$ is the residual energy of a sensor node i at time t and is defined as the remaining energy ($R_i(t)$) of node i to the initial energy (\hat{E}) of each node at time t . For all the nodes in the network, \hat{E} is similar at the time of deployment. For the second KPI, we consider the load fairness at the edge servers. $I_{n,i}$ is a binary indicator, i.e., $I_{n,i}$ is 1, if an n th gateway transmits φ_n packets to i th server at time t , otherwise, $I_{n,i}$ is 0.

To find an optimal Gateway-Edge configuration, an SDN controller needs to perform an exhaustive search for all possible gateway to edge combinations. Literally, it means that the size of the search space is equivalent to M^N , where M represents the number of edge servers and N represents the number of active gateways at a particular time t . The number of possible configurations increases exponentially with an increase in the number of M and N , respectively. To resolve the Gateway-Edge configuration as an optimization problem, we use the evolutionary algorithms, i.e., GA and DPSO. The following steps are executed for these algorithms to achieve an optimal configuration and a balanced load.

1. Generate a random population R^0 of size Δ . The best possible position for each particle, i.e., Gateway, is initiated such that $Pbest_i^0 = r_i^0, \forall 1 \leq i \leq \Delta$.
2. Discover the fitness value of each particle for DPSO and each chromosome for GA in R^0 (using Eq.(5)) and identify its global best position $Gbest^0$, using Eq.(6).

$$Gbest^0 = argmax_{1 \leq i \leq \Delta} F(Pbest_i^0).$$

3. For GA, if the best candidate solution for Gateway-Edge configuration is attained or the maximum number of generations has reached, then the search ends, otherwise, Step 4 is executed. For DPSO, if the best candidate solution is achieved, then the velocities of particles in the current population need to be updated using Eq.(7).

$$v_i^I = j_w v_i^{I-1} + a_1 r_1 (Pbest_i^I - x_i^I) + a_2 r_2 (Gbest_i^I - x_i^I). \quad (7)$$

Here, x_i^I represents the current position of particle i at I^{th} iteration, r_1 and r_2 are random variables within the (0, 1) range, a_1 and a_2 are acceleration constants used for pulling the particles toward the best position, and j_w reflects the inertia effect of preceding particle's velocity over the updated particle's velocity.

4. Next, a set of the best available γ chromosomes are extracted from the current population for GA. The current population is R^I and the selection probability is P_g . For DPSO, the iteration number is simply updated, i.e., $I=I+1$.
5. In case of GA, crossover and mutation are performed on γ . All infeasible solutions, i.e., $R^I - \gamma$ are replaced with μ . Here, μ represents the newly generated chromosomes. In the case of DPSO, if the best candidate solution is attained for Gateway-Edge configuration, then the search ends; otherwise, Step 6 is executed.
6. For GA, all the steps from Step 2 are repeated. For DPSO, the personal best position for each particle is updated using Eq.(8).

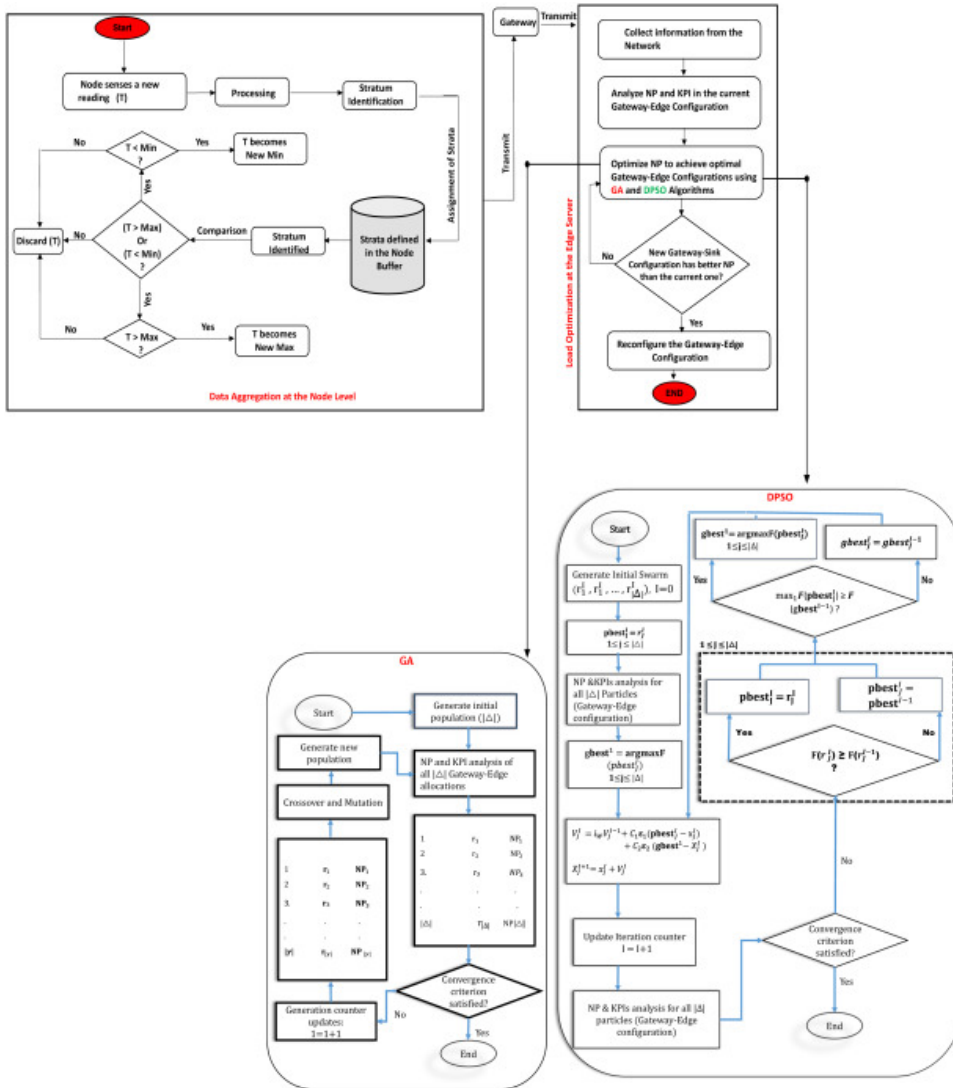
$$Pbest_i^I = \begin{cases} Pbest_i^{I-1}, & \text{if } F(r_i^I) \leq F(Pbest_i^{I-1}) \\ r_i^I, & \text{otherwise.} \end{cases} \quad (8)$$

7. For DPSO, the global best position is updated using Eq.(9).

$$Gbest_i^I = \begin{cases} \operatorname{argmax}_{1 \leq i \leq \Delta} F(Pbest_i^I), & \text{if } F(Pbest_i^I) > F(Pbest_i^{I-1}) \\ Gbest_i^{I-1}, & \text{otherwise.} \end{cases} \quad (9)$$

8. For DPSO, repeat all the steps from Step 1.

The flowchart of our proposed approach is shown in Fig.3. The SDN controller constantly performs the Gateway-Edge configuration based on the NP and KPI values. Since the controller needs to collect various information from the network to analyse the NP for the current Gateway-Edge configuration, we have highlighted the data fusion at the node level as well. The above solution can be used for homogeneous edge servers or that have capabilities to execute at approximate equal times. Furthermore, it does not account for dynamic scenarios where some data get processed earlier than the other edges. Therefore, to further balance the load, a service migration algorithm is suggested – which is feasible as, largely, edge servers run Linux-based operating systems.



Download : [Download high-res image \(693KB\)](#)

Download : [Download full-size image](#)

Fig. 3. Flowchart of the proposed approach.

3.3.1. Service migration technique

In our framework, load balancing can bring, at least, two benefits: performance improvement; and infrastructure energy efficiency. We trigger service migration either if: the utilization level of a particular edge node; and/or the data transfer rate on a particular link (channel condition), exceed certain pre-defined threshold values (steps 1 to 5)[36]. Once a module is being moved to another edge, it will immediately start receiving packets on another, perhaps, less utilized route. This could be achieved, after copying memory contents of the VM or container through sending a complimentary ARP (address resolution protocol) reply packet to inform the routing devices, within the network, to send data packets to its new location. As a result, both goals, i.e. balanced workload on various edges and reduced network traffic, could be achieved. Once a migration decision is triggered from a particular edge, next is to select a module of a suitable application to move. We move the application's module which can utilize the destination edge more i.e. priority is given to the module which is receiving more packets than other modules (step 6). Lastly, the module is migrated to the least utilized, neighbouring, edge platform; in order to diminish the migration performance impacts over the application's module and migration time (step 7). Finally, the selected module of the application m and the destination server are added to the migration map (step 8–10). The migration map is, then, passed to the load optimization module, as shown in Fig. 3, in order to reconfigure the Gateway-Edge



configuration, periodically. The migration steps are described in Algorithm 1:

Algorithm 1: Service migration technique

Input: dynamic CPU and memory utilization threshold values i.e. U_{tc} and U_{tm} , respectively - computed periodically using Eq. (12); channel condition C_t

Output: migration list $map \rightarrow$ input for load optimization module in Fig. 3

```

1 compute CPU utilization level of the edge node ( $EN_{uc}$ );
2 compute memory utilization level of the edge node ( $EN_{um}$ );
3 compute channel condition ( $C_c$ );
4 for each node  $\in$  edge do
5   if  $EN_{uc} \geq U_{tc}$  or  $EN_{um} \geq U_{tm}$  or  $C_c \geq C_t$  then
6     select application  $m$  from edge node;
7     choose edge node  $n$  as destination node;
8      $map \leftarrow m, n$ ;
9   end if
10 end for
11 return  $map$ 

```

[Download : Download high-res image \(186KB\)](#)

[Download : Download full-size image](#)

where EN_{uc} is the total provisioned CPU resources (cores) and EN_{um} is the total provisioned memory resources (RAM) with respect to their total capacities. Note that, U_{tc} and U_{tm} refer to the utilization level of a particular resource i.e. CPU, memory, respectively. Network resources such as bandwidth can also be considered in this formulation. Moreover, the channel condition B_{ij} is estimated using the transmission rate T_r , as given by:

$$T_r = B_{ij} \cdot \log_2 \left(1 + \frac{P_{ij} \cdot h_{ij}}{N} \right) \quad (10)$$

where B_{ij} represents the bandwidth between edge server i and gateway j , h_{ij} denotes the channel gain for gateway j at edge server i and P_{ij} is the transmission power of gateway j . Furthermore, N is the background noise [37]. Note that, Alg. 1 will approximately take $\mathcal{O}(mn \log(n))$ - where m denotes the total number of edges, n denotes the number of edge nodes and $\log(n)$ is the time needed to compute configuration states such as resource utilization levels and channel conditions. The best case occurs at $\mathcal{O}(\log(n))$ plus the time needed to complete all possible migrations. However, complexity would increase up to $\mathcal{O}(mn)^2$ for large number of edges, hosts and application requests – if unluckily an application cannot be placed or, in case, enough resources are not available. Note that, from security point of view, service migration in the IoT and VM or container migration in infrastructure clouds are completely different [36]. Usually, in infrastructure clouds, the migration data is transferred over dedicated networks; however, in IoT the data is transferred over the internet. This makes it essential to encrypt the migrated data and to authenticate the service migration messages that are exchanged among various edge devices.

Using static values for thresholds may not be feasible to trigger effective migrations in platforms with dynamic, heterogeneous and unpredictable workloads. This is due to the fact that resources that falls within the range of the least and most utilized (lower and upper thresholds) resources could not be reconfigured i.e. all hosts are equally loaded. In such scenario, threshold values can either be decreased or increased to balance the load amongst the edge nodes. Therefore, threshold values are needed to be adaptive and dynamically estimated using some sort of statistical techniques on historical data [23]. For example, we can adjust the threshold values based on the strength of the deviation of the edge or link utilization levels because higher deviations increase the likelihood of rising utilization levels. In other words, the higher the deviation, the lower the value of the threshold. Various methods such as local regression (LR), median absolute deviations (MAD), and entropy can be used to measure the statistical dispersion. For implementational simplification purposes, we prefer to use the MAD that describes the median of absolute values of deviations (residuals) from the data's median. For a particular dataset

$\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \dots, \mathcal{D}_n\}$, the MAD can be computed as:

$$MAD = \text{median}_i (\text{abs} [\mathcal{D}_i - \text{median}_j (\mathcal{D}_j)]) \quad (11)$$

The adaptive threshold value \mathcal{T}_v is given by:

$$\mathcal{T}_v = 1 - \lambda \cdot MAD \quad (12)$$

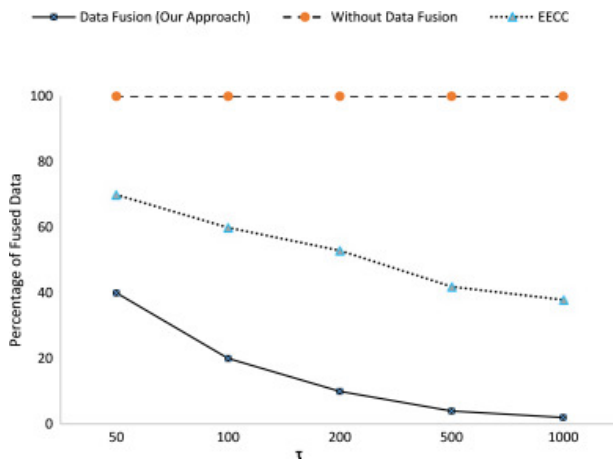
where λ is a parameter that describes how strongly the system tolerates edges over utilizations – lower λ results in higher tolerance to variations in utilization level. Once \mathcal{F}_v is computed, the utilization levels of the node and link are compared to it in order to trigger appropriate migration decisions.

4. Experimental results

In this section, we evaluate the efficiency of our proposed data fusion and load optimization approach in terms of various experimental metrics. For data fusion, we developed a Java-based simulator that utilizes the data collected from sensors, a setup similar to the one adopted at Intel Berkeley Research Lab[38]. Upon fusion, the simulator feeds the refined data to the gateways. For optimal Gateway-Edge configuration, we use Matlab 2018a interfaced with Java. Moreover, we added several Java class files to mimic the notion of containers that simulate a containerized fog infrastructure. The classes were taken from the well-known fog simulator iFogSim[20]. The service migration technique uses either: (i) a static threshold value of 80%; or (ii) dynamic thresholds computed using Eq.(12) in order to trigger migrations of application modules across edges. We further assume that overload (i.e. upper threshold) will not happen due to service placement constraint. To carry out this, the iFogSim default policies for selecting over-loaded servers, containers and target servers were used.

In Fig.4, the percentage of fused packets transmitted to the gateways is shown for different values of τ . Here, τ represents the number of readings sensed by each node over its sampling interval (S_r). The percentage of transmitted packets is calculated as $\frac{S_p}{\tau} \times 100$, where S_p denotes the number of fused packets sent from the strata of each node. The efficiency of our data fusion approach enhances with an increase in the value of τ . The percentage of transmitted fused packets from the strata of each node drops to 1% for 1000 packets, sensed during S_r . Our approach conserves the energy of resource-starving nodes and at the same time, reduces the burden on the network gateways. In comparison to our approach, the existing schemes deliver higher percentage of redundant data to the gateways. For example, EECC[34] transmits multiple copies of the same data from the strata of each node after S_r interval. As a result, the percentage of fused data delivered at the gateways is proportionally high. Moreover, without data fusion, all the sensed packets need to be transmitted to the gateways that will adversely affect the decision-making at the data centres.

During data fusion, each node examines the similarity index (Ω) in the data gathered over the S_r interval. This index further reduces the redundancy and at the same time, lowers the processing burden on the nodes and the network gateways. In Fig.5, the percentage of fused data for varying values of Ω is shown. In this figure, the values of τ varies from 200 to 1000 and Ω from 0.03 to 0.1, respectively. If Ω is 0.03, it means that among multiple readings having a similarity lower than or equal to 0.03, only one reading will be retained and the rest will be discarded. As a result, a higher percentage of readings will be discarded with an increase in the value of Ω . Moreover, our approach achieves a higher percentage of fusion when the value of τ increases. This figure shows that with higher values of Ω and τ , the processing and transmission burdens on the edge nodes and gateways decreases, significantly. In the absence of data fusion technique, a higher percentage of data is delivered to the gateways that in turn increases the processing and transmission burden on the nodes.

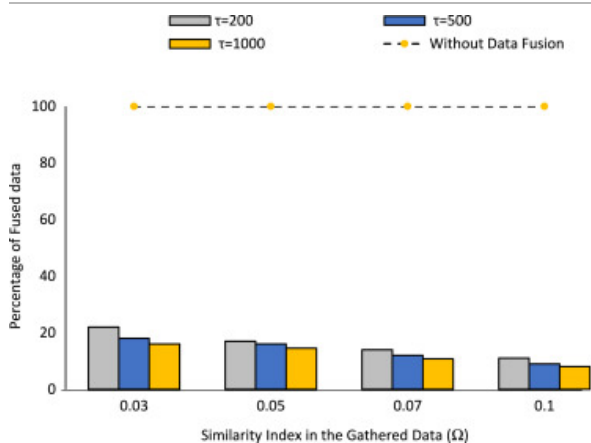


[Download : Download high-res image \(113KB\)](#)

[Download : Download full-size image](#)

Fig. 4. Percentage of fused data.

The optimal Gateway-Edge configurations achieved by GA and DPSO are shown in Table 1. We considered three benchmark problems P_1 , P_2 , and P_3 with two, three, and four edge servers, respectively. P_1 , P_2 , and P_3 contain 1200 sensor nodes including 200 gateways, distributed over the sensing field. In P_1 , GA converges to an optimal Gateway-Edge configuration after the 14th generation. DPSO, on the other hand, achieves the optimal configuration after the 11th iteration. Please note that iteration and generation are similar terms. The former is used in PSO and the latter in GA, as discussed in Section 3.3. GA achieves 185 optimal Gateway-Edge configurations over 200 generations with a convergence rate of 0.925, whereas, in DPSO, there are 188 optimal configurations with a convergence rate of 0.94. In P_2 and P_3 , the convergence rate of GA and DPSO decreases and larger values of iterations and generations are required to achieve an optimal Gateway-Edge configuration. It is mainly due to an increasing number of edge servers in these benchmark problems. These results show that DPSO reaches an optimal solution in fewer iterations as compared to GA.



[Download : Download high-res image \(102KB\)](#)

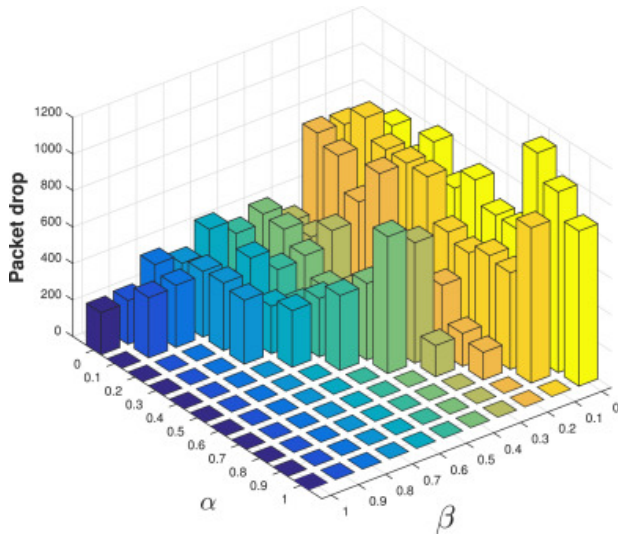
[Download : Download full-size image](#)

Fig. 5. Data fusion with varying values of similarity index.

We assessed the network performance (NP) for all optimal solutions in term of packet drop by modifying the KPIs such that α and β are set in the ordered form of 0,0.1, ...,1 with a constraint $\alpha + \beta \leq 1$. To properly tune α and β , an exhaustive search is performed on P_1 using these parameters to achieve an optimal Gateway-Edge configuration. When α and β are set to 0.8 and 0.2, respectively; a minimum packet drop is observed, as shown in Fig.6. The selection of proper weights for the optimization function, i.e., NP, is challenging and essential for achieving optimal results in the context of evolutionary algorithms.

PDF

Help



[Download : Download high-res image \(467KB\)](#)

[Download : Download full-size image](#)

Fig. 6. Packet loss with varying values of α and β .

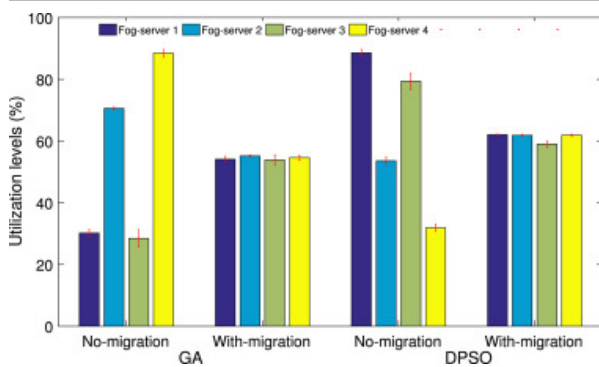
Table 1. Optimal gateway-edge configurations.

Number of edge servers	Convergence rate		Number of iterations	
	GA	DPSO	GA	DPSO
2	0.925	0.94	14	11
3	0.86	0.895	28	21
4	0.805	0.85	43	31

The service migration technique, as suggested in Section 3.3.1, was implemented to balance the load across the edge servers. Increasing the total number of edge servers decreases the utilization levels and vice versa, as shown in Table 2. Moreover, the number of migrations happened is proportional to the amount of fog servers. The standard deviation actually represents how the current load on each server differs from other servers – the higher this value, the more less balanced is the workload and vice versa. We observed significant reduction in utilization levels, that, essentially translate to greater energy savings and improved levels of performance. We observed, as shown in Table 2, that using static threshold values reduces the migration opportunities. Moreover, increased levels of variations were observed in server's utilization levels. This means that either resources were utilized more or the least due to less migration opportunities. Using dynamic threshold values, the variations in utilization levels may decrease significantly – as more migrations will occur subsequently. Varying the threshold values will essentially result in variations of outcomes and differences in Gateway-Edge reconfiguration. Fig. 7 shows average utilization levels (along with error bars at five minute intervals) when four fog servers are taken into account. Moreover, number of migrations would have some impacts on network traffic and processing performance. For four fog servers when migrations are not taken into account, we achieved 48.75 ± 23.67 and 41.28 ± 28.56 average utilization levels for GA and DPSO, respectively. The high variations (standard deviations) show the imbalance load across various servers which might happened due to dynamics in workloads execution or processing patterns. With migrations, we were able to significantly reduce these variations as shown in Fig. 7.

Table 2. Average utilization levels (%) of the edge servers and number of migrations (using static and dynamic threshold values for load balancing and resource re-configuration).

Number of servers	GA		DPSO	
	Utilization	Migrations	Utilization	Migrations
Static threshold values				
2	62.9±5.99	101	61.4±6.45	89
3	60.4±6.31	105	58.7±8.01	103
4	59.6±6.44	147	59.1±7.32	122
Dynamic threshold values				
2	71.2±1.27	134	75.3±2.02	155
3	62.8±1.71	178	68.4±3.89	189
4	54.3±1.09	201	59.7±2.99	217



[Download : Download high-res image \(155KB\)](#)

[Download : Download full-size image](#)

Fig. 7. Load balancing for four fog servers [error bars denote standard deviations from the means].

5. Conclusions and future work

In this paper, we proposed a lightweight data fusion and AI-enabled load optimization approach for reconfigurable IoT applications. The buffer of each node is partitioned into strata that hold and transmit only non-correlated fused data towards the network gateways and edge servers. We used GA and DPSO to optimize the usage of available resources by identifying the optimal routes for upstream transmission of refined data from the gateways to edge servers. These algorithms monitored the load at the servers, and if an unbalanced load is experienced, the current Gateway-Edge configuration is reconfigured. For load monitoring at the edge, various Key Performance Indicators (KPIs) were used. Our experimental results significantly reduced the processing and transmission burden at the nodes for large-sized data streams. Our approach achieved optimal gateway-edge configurations for varying number of edge servers in a densely populated network setup. Moreover, a migration approach was used to balance the load across different edge servers. Our evaluation of the proposed migration approach demonstrated that all edge servers are relatively utilized uniformly while having lower standard deviations in their utilization levels. Subsequently, this ensures that data is processed at edge which increases performance. In the future, we aim to analyse the network performance by maintaining a balanced load at the network gateways. It will enable the gateways to automate the downstream transmission links towards the nodes. Moreover, we are keen to see the impact of migrations in dynamic scenarios, particularly, on network traffic and transmission delays.

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.

Acknowledgements

This work was partially supported by a pilot award from the Center for Research in Human Movement Variability, USA and the NIH, USA (P20GM109090) and the National Science Foundation, USA under award CNS-2016714.

[Special issue articles](#) [Recommended articles](#)

References

- [1] Shen Y., Zhang T., Wang Y., Wang H., Jiang X.
Microthings: A generic IoT Architecture for flexible data aggregation and scalable service cooperation
IEEE Commun. Mag., 55 (9) (2017), pp. 86-93
[View in Scopus](#) ↗ [Google Scholar](#) ↗
- [2] Khan F., Jan M.A., Rehman A.U., Mastorakis S., Alazab M., Watters P.
A secured and intelligent communication scheme for IIoT-enabled pervasive edge computing
IEEE Trans. Ind. Inf. (2020)
[Google Scholar](#) ↗
- [3] Ding W., Jing X., Yan Z., Yang L.T.
A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion
Inf. Fusion, 51 (2019), pp. 129-144
[View PDF](#) [View article](#) [View in Scopus](#) ↗ [Google Scholar](#) ↗
- [4] Alam F., Mehmood R., Katib I., Albogami N.N., Albeshri A.
Data fusion and IoT for smart ubiquitous environments: a survey
IEEE Access, 5 (2017), pp. 9533-9554
[View in Scopus](#) ↗ [Google Scholar](#) ↗
- [5] Twayej W., Khan M., Al-Raweshidy H.S.
Network performance evaluation of M2M with self organizing cluster head to sink mapping
IEEE Sens. J., 17 (15) (2017), pp. 4962-4974
[View in Scopus](#) ↗ [Google Scholar](#) ↗
- [6] Jan M.A., Khan F., Khan R., Mastorakis S., Menon V.G., Watters P., Alazab M.
A lightweight mutual authentication and privacy-preservation scheme for intelligent wearables devices in industrial-CPS
IEEE Trans. Ind. Inf. (2020)
[Google Scholar](#) ↗
- [7] Mazza D., Tarchi D., Corazza G.E.
A unified urban mobile cloud computing offloading mechanism for smart cities
IEEE Commun. Mag., 55 (3) (2017), pp. 30-37
[View in Scopus](#) ↗ [Google Scholar](#) ↗
- [8] El-Sayed H., Sankar S., Prasad M., Puthal D., Gupta A., Mohanty M., Lin C.-T.
Edge of things: The big picture on the integration of edge, IoT and the cloud in a distributed computing environment

IEEE Access, 6 (2017), pp. 1706-1717

[View in Scopus](#) ↗ [Google Scholar](#) ↗

- [9] Sharma S.K., Wang X.
Live data analytics with collaborative edge and cloud processing in wireless IoT networks
IEEE Access, 5 (2017), pp. 4621-4635

[View in Scopus](#) ↗ [Google Scholar](#) ↗

- [10] Yao H., Li M., Du J., Zhang P., Jiang C., Han Z.
Artificial intelligence for information-centric networks
IEEE Commun. Mag., 57 (6) (2019), pp. 47-53

[CrossRef](#) ↗ [View in Scopus](#) ↗ [Google Scholar](#) ↗

- [11] Al-Janabi T.A., Al-Raweshidy H.S.
A centralized routing protocol with a scheduled mobile sink-based AI for large scale I-Iot
IEEE Sens. J., 18 (24) (2018), pp. 10248-10261

[CrossRef](#) ↗ [View in Scopus](#) ↗ [Google Scholar](#) ↗

- [12] Bijarbooneh F.H., Du W., Ngai E.C.-H., Fu X., Liu J.
Cloud-assisted data fusion and sensor selection for internet of things
IEEE Internet Things J., 3 (3) (2016), pp. 257-268

[View in Scopus](#) ↗ [Google Scholar](#) ↗

- [13] Nesa N., Banerjee I.
Sensorrank: An energy efficient sensor activation algorithm for sensor data fusion in wireless networks
IEEE Internet Things J. (2018)

[Google Scholar](#) ↗

- [14] Kumar S., Chaurasiya V.K.
A strategy for elimination of data redundancy in internet of things (IoT) based wireless sensor network (WSN)
IEEE Syst. J. (2018)

[Google Scholar](#) ↗

- [15] Sanyal S., Zhang P.
Improving quality of data: IoT data aggregation using device to device communications
IEEE Access, 6 (2018), pp. 67830-67840

[CrossRef](#) ↗ [Google Scholar](#) ↗

- [16] Vose M.D.
The Simple Genetic Algorithm: Foundations and Theory, Vol. 12
MIT press (1999)

[Google Scholar](#) ↗

- [17] Trelea I.C.
The particle swarm optimization algorithm: convergence analysis and parameter selection
Inf. Process. Lett., 85 (6) (2003), pp. 317-325

 [View PDF](#) [View article](#) [View in Scopus](#) ↗ [Google Scholar](#) ↗

- [18] Abbasi M., Rafiee M., Khosravi M.R., Jolfaei A., Menon V.G., Koushyar J.M.
An efficient parallel genetic algorithm solution for vehicle routing problem in cloud implementation of the intelligent transportation systems

J. Cloud Comput., 9 (1) (2020), p. 6

[View in Scopus](#) ↗ [Google Scholar](#) ↗

[19] Kaur K., Garg S., Kaddoum G., Ahmed S.H., Atiquzzaman M.

KEIDS: Kubernetes-based energy and interference driven scheduler for industrial IoT in edge-cloud ecosystem

IEEE Internet Things J., 7 (5) (2020), pp. 4228-4237

[CrossRef](#) ↗ [View in Scopus](#) ↗ [Google Scholar](#) ↗

[20] Gupta H., Vahid Dastjerdi A., Ghosh S.K., Buyya R.

iFogSim: A toolkit for modeling and simulation of resource management techniques in the Internet of Things, Edge and Fog computing environments

Softw. - Pract. Exp., 47 (9) (2017), pp. 1275-1296

[CrossRef](#) ↗ [View in Scopus](#) ↗ [Google Scholar](#) ↗

[21] Bittencourt L.F., Diaz-Montes J., Buyya R., Rana O.F., Parashar M.

Mobility-aware application scheduling in fog computing

IEEE Cloud Comput., 4 (2) (2017), pp. 26-35

[View in Scopus](#) ↗ [Google Scholar](#) ↗

[22] Machen A., Wang S., Leung K.K., Ko B.J., Salonidis T.

Live service migration in mobile edge clouds

IEEE Wirel. Commun., 25 (1) (2018), pp. 140-147

[CrossRef](#) ↗ [View in Scopus](#) ↗ [Google Scholar](#) ↗

[23] Beloglazov A., Buyya R.

Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints

IEEE Trans. Parallel Distrib. Syst., 24 (7) (2013), pp. 1366-1379, [10.1109/TPDS.2012.240](#) ↗

[View in Scopus](#) ↗ [Google Scholar](#) ↗

[24] Khan A.A., Zakarya M., Khan R.

H² – a hybrid heterogeneity aware resource orchestrator for cloud platforms

IEEE Syst. J., 13 (4) (2019), pp. 3873-3876

[CrossRef](#) ↗ [View in Scopus](#) ↗ [Google Scholar](#) ↗

[25] Zakarya M., Gillam L.

Managing energy, performance and cost in large scale heterogeneous datacenters using migrations

Future Gener. Comput. Syst., 93 (2019), pp. 529-547

 [View PDF](#) [View article](#) [View in Scopus](#) ↗ [Google Scholar](#) ↗

[26] Zakarya M., Gillam L., Ali H., Rahman I., Salah K., Khan R., Rana O., Buyya R.

EpcAware: A game-based, energy, performance and cost efficient resource management technique for multi-access edge computing

IEEE Trans. Serv. Comput. (2020), pp. 1-14, [10.1109/TSC.2020.3005347](#) ↗

[View in Scopus](#) ↗ [Google Scholar](#) ↗

[27] Khan A.A., Zakarya M., Buyya R., Khan R., Khan M., Rana O.

An energy and performance aware consolidation technique for containerized datacenters

IEEE Trans. Cloud Comput. (2019), [10.1109/TCC.2019.2920914](#) ↗

[Google Scholar](#) ↗

PDF

Help

- [28] Zakarya M., Gillam L., Khan A.A., Rahman I.U.
Perficientcloudsim: a tool to simulate large-scale computation in heterogeneous clouds
J. Supercomput. (2020), pp. 1-55, [10.1007/s11227-020-03425-5](https://doi.org/10.1007/s11227-020-03425-5) [↗](#)
[View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [29] Mastorakis S., Mtibaa A., Lee J., Misra S.
Icedge: When edge computing meets information-centric networking
IEEE Internet Things J., 7 (5) (2020), pp. 4203-4217
[CrossRef](#) [↗](#) [View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [30] Nour B., Mastorakis S., Mtibaa A.
Compute-less networking: Perspectives, challenges, and opportunities
IEEE Netw., 34 (6) (2020), pp. 259-265
[CrossRef](#) [↗](#) [View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [31] Mastorakis S., Mtibaa A.
Towards service discovery and invocation in data-centric edge networks
2019 IEEE 27th International Conference on Network Protocols (ICNP), IEEE (2019), pp. 1-6
[CrossRef](#) [↗](#) [Google Scholar](#) [↗](#)
- [32] Wang Q., Lin D., Yang P., Zhang Z.
An energy-efficient compressive sensing-based clustering routing protocol for WSNs
IEEE Sens. J., 19 (10) (2019), pp. 3950-3960
[View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [33] Li T., Bolic M., Djuric P.M.
Resampling methods for particle filtering: classification, implementation, and strategies
IEEE Signal Process. Mag., 32 (3) (2015), pp. 70-86
[View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [34] Jan S.R.U., Jan M.A., Khan R., Ullah H., Alam M., Usman M.
An energy-efficient and congestion control data-driven approach for cluster-based sensor network
Mob. Netw. Appl., 24 (4) (2019), pp. 1295-1305
[CrossRef](#) [↗](#) [View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [35] Jain R., Durresi A., Babic G.
Throughput fairness index: An explanation
ATM Forum Contribution, Vol. 99 (1999)
[Google Scholar](#) [↗](#)
- [36] Mach P., Becvar Z.
Mobile edge computing: A survey on architecture and computation offloading
IEEE Commun. Surv. Tutor. (2017)
[Google Scholar](#) [↗](#)
- [37] Zhang H., Guo F., Ji H., Zhu C.
Combinational auction-based service provider selection in mobile edge computing networks
IEEE Access, 5 (2017), pp. 13455-13464
[View in Scopus](#) [↗](#) [Google Scholar](#) [↗](#)
- [38] Madden S.

Intel berkeley research lab data

(2003)

[Google Scholar](#) ↗

Cited by (24)

AI augmented Edge and Fog computing: Trends and challenges

2023, Journal of Network and Computer Applications

[Show abstract](#) ✓

Energy-efficient secure data fusion scheme for IoT based healthcare system

2023, Future Generation Computer Systems

[Show abstract](#) ✓

An IoT-based resource utilization framework using data fusion for smart environments

2023, Internet of Things (Netherlands)

Citation Excerpt :

...This includes consuming more resources and processing time for the massive and redundant data, downgrading resource utilization accuracy as well as availability and reliability due to the faulty, real-time, and heterogenous IoT data. In such cases, data fusion appears to be a solution for handling IoT data features that would implicitly resolve the resource utilization concerns in IoT-based systems [18]. Data fusion is the process of handling data from multiple data sources to produce more consistent, accurate, and useful information than those provided by any individual data source [19]...

[Show abstract](#) ✓

Smart-3DM: Data-driven decision making using smart edge computing in hetero-crowdsensing environment

2022, Future Generation Computer Systems

Citation Excerpt :

...Similarly, authors in [14] develop a Virtualized Network Function (VNFs) edge cloud architecture that considers the dynamicity of the network changes and the latency-sensitivity of the applications while automatically allocate the resources in the edge and the cloud. Also, the authors in [15] consider the application with low time tolerance and high delay-sensitivity where they tried to improve the application QoS through proposing a lightweight data fusion optimization. The proposed mechanism relies on eliminating the correlation and redundancy of the sensed data and adopting a dynamic migration technique that allows reducing the overhead while maintaining a periodic re-configuration of the ES...

[Show abstract](#) ✓

Multi-tier delay-aware load balancing strategy for 5G HC-RAN architecture

2022, Computer Communications

[Show abstract](#) ✓

The central role of data repositories and data models in Data Science and Advanced Analytics

2022, Future Generation Computer Systems

[Show abstract](#) ✓

PDF

Help

[↗](#) View all citing articles on Scopus



Mian Ahmad Jan is an Assistant Professor at the Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan. He received his Ph.D. in Computer Systems from the Faculty of Engineering and Information Technology (FEIT), University of Technology Sydney (UTS), Australia. His research interests include energy-efficient and secured communication in Wireless Sensor Networks and Internet of Things. He has been guest editor of numerous special issues in various prestigious journals e.g. Future Generation Computer Systems, IEEE Transactions on Industrial Informatics, Springer Neural Networks and Applications, IEEE Sensor, etc. are few to mention. He is an IEEE senior member.



Muhammad Zakarya received the Ph.D. degree in Computer Science from the University of Surrey, Guildford, U.K. He is currently a Lecturer with the Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan. His research interests include cloud computing, mobile edge clouds, Internet of Things (IoT), performance, energy efficiency, algorithms, and resource management. He has deep understanding of the theoretical computer science and data analysis. Furthermore, he also owns deep understanding of various statistical techniques which are, largely, used in applied research. He is an Associate Editor for the IEEE Access Journal and has served as TPC member in various international conferences and workshops.



Muhammad Khan received his Ph.D. degree in wireless communications from Brunel University London, United Kingdom. He is associated with the Wireless Network and Communication Centre (WNCC) at Brunel University London, and the Computer Networks (ComNets) Lab at New York University (NYU). He is currently working on Projects related to 5G mmWave Congestion Control in collaboration with the University of Contabria, Spain and University of Limerick, Ireland. His main research interest is next generation wireless communications, Cloud Radio Access networks, Artificial Intelligence, Machine learning and Congestion Control.



Spyridon Mastorakis (smastorakis@unomaha.edu) is an Assistant Professor in Computer Science at the University of Nebraska Omaha. He received his Ph.D. in Computer Science from the University of California, Los Angeles (UCLA) in 2019. He also received an M.S. in Computer Science from UCLA in 2017 and a 5-year diploma (equivalent to M.Eng.) in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 2014. His research interests include network systems and protocols, Internet architectures (such as Information-Centric Networking and Named-Data Networking), edge computing and IoT, and security.



Varun G. Menon is currently an Associate Professor in Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. He is a Senior Member of IEEE and a Distinguished Speaker of ACM Distinguished Speaker. Dr. Varun G Menon is currently a Guest Editor for IEEE Transactions on Industrial Informatics, IEEE Sensors Journal, IEEE Internet of Things Magazine and Journal of Supercomputing. He is an Associate Editor of IET Quantum Communications and also an Editorial Board Member of IEEE Future Directions: Technology Policy and Ethics. His research interests include Internet of Things, Fog Computing and Networking, Underwater Acoustic Sensor Networks, Cyber Psychology, Hijacked Journals, Ad-Hoc Networks, Wireless Sensor Networks.

PDF

Help



Venki Balasubramanian received the Ph.D. degree in body area wireless sensor network (BAWSN) for remote healthcare monitoring applications. He is the Pioneer in building (pilot) remote healthcare monitoring application (rHMA) for pregnant women for the New South Wales Healthcare Department. His research establishes a dependability measure to evaluate rHMA that uses BAWSN. His research opens up a new research area in measuring time-critical applications. He contributed immensely to eResearch software research and development that uses cloud-based infrastructure and a core member for the project sponsored by Nectar Australian research cloud provider. He contributed heavily in the field of healthcare informatics, sensor networks, and cloud computing. He also founded Anidra Tech Ventures Pty Ltd a smart remote patient monitoring company.



Ateeq Ur Rehman is currently working as a Lecturer at the department of Computer Science, Abdul Wali Khan University Mardan, KPK, Pakistan. He received his Ph.D. degree from the University of Southampton in 2017. His main research interests are next-generation wireless communications and cognitive radio networks, Internet of Things, Internet of Vehicles, blockchain technology and differential privacy.

1 https://criu.org/Main_Page ↗.

2 Plural of stratum.

[View Abstract](#)

© 2021 Elsevier B.V. All rights reserved.



Copyright © 2023 Elsevier B.V. or its licensors or contributors.
ScienceDirect® is a registered trademark of Elsevier B.V.



PDF

Help

Detection and robustness evaluation of android malware classifiers

Original Paper | Published: 26 June 2021

Volume 18, pages 147–170, (2022) [Cite this article](#)



[Journal of Computer Virology and
Hacking Techniques](#)

[Aims and scope](#) →

[Submit manuscript](#) →

Abstract

Android malware attacks are tremendously increasing, and evasion techniques become more and more effective. For this reason, it is necessary to continuously improve the detection performances. With this paper, we wish to pursue this purpose with two contributions. On one hand, we aim at evaluating how improving machine learning-based malware detectors, and on the other hand, we investigate to which extent adversarial attacks can deteriorate the performances of the classifiers. Analysis of malware samples is performed using static and dynamic analysis. This paper proposes a framework for integrating both static and dynamic features trained on machine learning methods and deep neural network. On employing machine learning algorithms, we obtain an accuracy of 97.59% with static features using SVM, and 95.64% is reached with dynamic features using Random forest. Additionally, a 100% accuracy was obtained with CART and SVM using hybrid attributes (on combining relevant static and dynamic features). Further, using deep neural network models, experimental results showed an accuracy of 99.28% using static features, 94.61% using dynamic attributes, and 99.59% by combining both static and dynamic features (also known as multi-modal attributes). Besides, we evaluated the robustness of classifiers against evasion and poisoning attack. In particular comprehensive analysis was performed using permission, APIs, app components and system calls (especially *n*-grams of system calls). We noticed that the performances of the classifiers

significantly dropped while simulating evasion attack using static features, and in some cases 100% of adversarial examples were wrongly labelled by the classification models. Additionally, we show that models trained using dynamic features are also vulnerable to attack, however they exhibit more resilience than a classifier built on static features.

Authors and Affiliations

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

M. A. Arya, Josna Philomina, Anson Pinheiro & K. S. Ajith

About this article



Check for updates

Cite this article

Anupama, M.L., Vinod, P., Visaggio, C.A. *et al.* Detection and robustness evaluation of android malware classifiers. *J Comput Virol Hack Tech* **18**, 147–170 (2022). <https://doi.org/10.1007/s11416-021-00390-2>

[Download citation](#)

Received

07 August 2020

Accepted

31 May 2021

Published

26 June 2021

Issue Date

September 2022

DOI

<https://doi.org/10.1007/s11416-021-00390-2>

**INTERNATIONAL RESEARCH JOURNAL OF SCIENCE
ENGINEERING AND TECHNOLOGY**



ISSN 2454-3195

An Internationally Indexed Peer Reviewed & Refereed Journal

WWW.RJSET.COM
www.isarasolutions.com

Published by iSaRa Solutions

IoT based Power Analyzer for an Automated Home

Susmi Jacob, Shilpa P C, and Binu John

Susmi Jacob is with Department of Computer Science, SCMS school of Engineering and Technology, Kerala

Email: susmijacob@scmsgroup.org

Email: shilpacc@scmsgroup.org Email: binujohn@scmsgroup.org

Abstract— As technology enhances, houses have become smarter and energy-efficient. Traditional switches are gradually being replaced with automated centralized switches with remotecontrols in modern homes. The inmates find it difficult to activate the traditional wall switches located throughout the residence when they are needed. Smartphones give a modern option for remote-controlled home automation. The main goal of this study is to design and construct a low-cost Internet Of Things (IoT) enabled energy monitoring system that can be benefited in different applications such as energy management in smart automated homes as well as electricity billing systems. We develop a power analyzer using an Arduino board and Current Transformer (CT) sensor, as well as gadgets that can be controlled remotely using an Android OS smartphone. At the beneficiary end, a Bluetooth module is connected to an Arduinoboard, while at the transmitter end, a GUI application running on a PDA sends ON/OFF bearings to the authority where homeequipment is installed. Through this technique, the pieces of equipment are turned ON/OFF by tapping the correct spots on the GUI. Additionally, we recorded the daily and monthly readings of power consumed to monitor the power consumption in a month and to give necessary warnings and suggestions to the consumer.

I. INTRODUCTION

Energy conservation is an urgent requirement of this era. The idea of resourceful equipment in assorted areas such as air conditioning, refrigeration, lighting, etc will lead to energy-efficient utilization of household devices. Energy auditing is an inevitable mechanism for analyzing the systematic energy utilization of equipment and devices. It provides a better way to control the use of electricity in case of excess usage and also helps the user to fix the inaccuracy in the electricity bill which may show excess amount sometimes [1].

Domestic electricity bill which presents surplus amount that causes disapproval for the users. By using a smart power analyzer system user monitors the energy utilization details at the equipment level and governs it rather than calculating the fixed monthly expenditure. This may also aid the user to restore the normal

appliances with energy-efficient and smart devices[2]. Critically, the checking power system can caution the user on startling overabundance utilization brought about by the improper working, absence of timely maintenance, and so forth. Further, energy management in the proper way leads to the better utilization of the resources, and thereby we can reduce the cost.

Thus the wastage of energy can be reduced to meet future needs by protecting the precious resource. Similarly, the cost estimation can be predicted for each industrial unit or home for better utilization of the energy. Moreover, this comparison and the analysis of cost estimation for each industrial unit or home will reduce the production cost which will result in the tremendous profit of the industry.

The importance of saving electricity attributes to the fact that electricity is generated from natural resources, which are limited and reducing as time goes. The unsustainable use of natural resources not only affects the balance of nature but also makes the planet completely unfit to live. Saving electricity can reduce its production, thereby it reduces the manpower and cost of production. Similarly by producing electricity from other means like coal accelerates pollution. Thus pollution can then be controlled by limiting the production of electricity by consuming electricity efficiently and effectively to save the planet.

A. Motivation

Energy consumed by each industrial unit or home can be measured and analyzed based on time stamps. Thus, the energy used by different industrial units or home can be compared and analyze which consume more energy. Hence,

the industrial units or homes which possess more energy consumption can be taken care and inference can be made that whether the industrial unit or home are actually needed this much energy or whether this consumption is due to any faults in machines. Also, it can control electronic devices from anywhere at any time using the internet or within a Wi-Fi connection. Thus wastage of electricity gets reduced by controlling devices at the correct time. The electricity meter stationed in the individual constructions displays the energy utilized by the buildings. There is an urgent necessity for a novel power analyzing system.

The wastage of energy can be reduced to save the future by protecting their precious resource. By reducing the wastage of electricity, the production of electricity can be decreased. Electricity is produced from natural resources such as water etc. So by limiting the production of electricity will lead to less exploitation of these natural resources. Similarly, electricity can be produced from other means like coal which increases pollution. Thus, pollution, manpower, and cost can be reduced by limiting the production of electricity which can be achieved by saving and using electricity in efficient and effective ways to save the planet.

B. Contribution

Internet of Things (IoT) is relied upon to achieve an enormous measure of progress in the field of pervasive computing. IoT-based energy the board framework can contribute a ton to the preservation of energy. It can control different electric devices in the home from anywhere at any time using an internet connection or Wi-Fi. Similarly, the cost estimation can be predicted for each industrial unit or home. And this comparison and analysis of cost estimation from each industrial unit or home will reduce the production cost, hence will result in tremendous profit to the industry.

An IoT Based Power Analyzer is a general-purpose real-time mobile application. It is used to measure the usage of power consumed by each industrial unit or home. It also estimates the cost of each industrial unit or home.

We propose the implementation of a smart plug, an energy observation system that provides real-time information on device-level energy consumption. An Arduino microcontroller board, an ENC28J60 LAN module, and a current electrical device detecting element are used in the suggested device. The present sensor technology employed is non-invasive. The device's computer software is written in Android and the statistics are stored in a server. The end product is a smart plug that uses the Arduino-android platform to monitor a distant device.

The next half essentially carries out the style of an IoT sensible Home System (IoTSHS) which gives the remote control to sensible home appliances via android mobile phones, similarly like PC/Laptop. The controller accustomed style the IoTSHS is Arduino Uno micro-controller. A temperature detector is provided to analyse the surrounding temperature and alert the users if the regulation of the fan speed is required. The designed IoTSHS will edge the total elements within the community by facilitating technologically improved remote dominance for the sensible home.

The organization of the paper is as follows. Section II summarises the literature survey.

II. RELATED WORK

We could find several substantiating facts which are beneficial for our research work. Literature survey have been done in the area of power analyzer as well as home automation.

A. Power Analyzer

A survey done on already existing literature on energy consumption and analysing reveals that a tremendous change have been reported in the implementation [5]. For the necessities of acknowledging energy-sparing and outflow decrease for the smart network, this paper presents an observing foundation of the energy, the board framework which is coordinated with flexibly energy. It governs, reads, and evaluate the processed readings, of energy creation, energy transportation, and energy utilization for the smart system. The design of the community energy system is split into the acquisition layer, transmission layer, and management layer. The

information correspondence between information securing devices and server embraces Ethernet and TCP/IP convention. The information correspondence between field instruments and information procurement gadgets receives the RS-485 interface and Modbus correspondence convention.

The advancement of an observing platform for a smart network can improve the administration level as well as diminish the energy utilization of the network energy framework. A few studies done on the possibilities of better energy utilization proposes various methods [6]. The users are going to be alerted when the electricity usage in their home exceeds the limits to avoid the wastage of energy consumption.

Various studies done on the data acquisition and control of energy utilization efficient method. Smart Home Energy Management Systems Based on Non-Intrusive Load Monitoring [7] proposed a unique system of good domestic energy management systems incorporating each approaches, in order that correct energy utilization watching and assuming communication with the smart home device at the same time achieved. The good parts directly management the appliances, whereas the essential controller coordinates the info assortment. The key point is that the competence of mechanically aligning the appliances to their corresponding sockets, reducing the need for manual initial setup. We assure that our smart framework, if sophisticated widely, will profit not solely separate households by reducing current bills.

A Server Agent is the term given to the innovative microcontroller used in this project. While the initial PC-based server is offline, this agent collects information from buyers. Once all knowledge area units have dropped, the Server Agent can turn down the PC-based server again. This procedure minimises the amount of energy used [4].

We conducted several research on An Efficient Home Energy Management Solution based on Automatic Meter Reading, with a specific focus on household energy consumption, and proposed a simple and

effective system for reducing power waste in a home. They designed a home energy management system (HEMS) that uses a simple energy management mechanism to make it easier to implement. It also makes use of AMR (Automatic Meter Reading) network-based power line communication (PLC). They installed HEMS in real customers' homes and validated the results, resulting in a significant energy conservation outcome that is vital for power reduction.[9].

GAPMR stands for "automated power metre reading system using GSM network." It is a system that consists of GSM digital power metres installed in the client unit and an electricity e-billing system at the energy supplier's end. The GSM digital electric metre (GPM) could be a single part IEC61036, a standard digital kWh electric metre with incorporated GSM electronic equipment that uses the GSM network to report power usage readings. The readings are send back to the energy supplier as short messaging system (SMS) via a wireless manager[8]. On the service provider's side, there's a degree e-billing system that's used to monitor all SMS metre readings, encrypt the charge value, update the information, and send charge notifications to its various customers by SMS, email, web portal, and letters. The effectiveness and efficiency of automatic meter reading

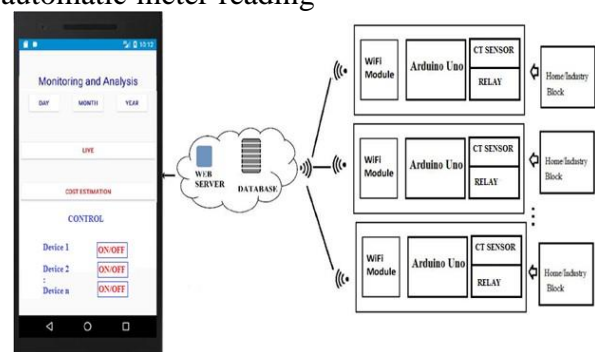


Fig. 1: Proposed system architecture

was demonstrated by implementing a GAPMR system. Also the charge and notification through the employment of GSM network can also be monitor by this system.

B. Home Automation

The research into the implementation of IoT for condition monitoring in homes shows that condition monitoring and energy management in the home may be done in an inexpensive, flexible, and cost-effective manner. The designed system's major tasks include remote control and management of home devices such as electrical lamps, heaters, and so on, as well as unassertive surveillance of domestic usage and supplying closed intelligence to reduce energy consumption using IoT technology.[3]. This will assist and schedule the individual's operation time according to the energy demand.

The majority of current sensible Web connection is provided by TV set-top boxes, allowing users to install and run additional advanced applications or plugins/add-ons for a certain platform. It means that a sensible—a wise TV set-top box is a good option for acting as a hub that integrates a variety of smart home solutions. This study presents a framework for managing household appliances that is enabled by a smart TV set-top box.

Many buttons (often dozens) are designed on the remote controller in home areas as the quality of devices/appliances improves, yet many of them are rarely used. A user is also perplexed by the controller, despite the fact that he or she only wants to do a simple task. This confusion in addition ends up in a far better likelihood of mal-functions. Additionally to the current a typical ways in which of communication between remote controllers and connected devices, like ventricose language (XML) messages, unit generally bandwidth-consumptive. The asymmetrical feature of Point-n-Press provides for simple and intuitive management by informing the target device and displaying the target's management interface on the remote controller's screen. Exclusively sensible pops that unit relevant to the present context unit by using the state dependencies of home device/equipment actions. Two real prototypes are being used to test the feasibility of the proposed theme. According to the findings, Point-n-Press could be a useful and relevant management theme for IoT-based smart homes.

III. PROPOSED SYSTEM

An IoT Based Power Analyzer with home automation is a real time IoT based mobile application which analyze the energy consumed by each industrial unit. It also predicts the cost estimation of each industrial unit or home which incorporates with online data storage. It clearly specifies the average energy used per day, per month and per year along with cost estimation. Energy consumed by different industrial units or home can be measured and analyzed based on time stamps. Thus, the energy used by different industrial units or home can be compare and analyze which consume more energy. So the industrial unit or home which possess more energy consumption can be taken care and inference can be made that whether the industrial units or home are actually needed this much energy or these much consumption

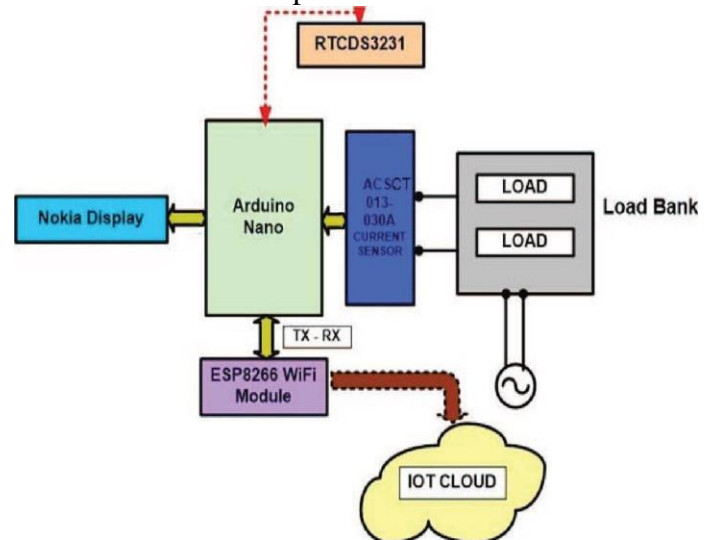


Fig. 2: Power Analyzer Architecture

is due to any faults in machines. By identifying faulty devices before their life expired, the production cost can be reduced to some extent because the faulty devices can be repaired before it get fully damaged.

The second section focuses on the design of an IoT sensible House System (IoTSHS), which might provide access to a smart home via a mobile device, similar to a PC or laptop. The Arduino Uno microcontroller is the standard controller for the IoTSHS. A temperature detector

is included to indicate the temperature in the area and inform the user if the fan speed needs to be adjusted. By giving improved remote dominance for the sensible house, the developed IoTSHS will edge the total elements within society.

A. Product Perspective

This app can control electronic equipment and to measure the usage or consumption of electricity for each device. Cost of consumption per devices can also available in this app. Thus the electricity usage can be analyzed and provides an efficient way electricity usage. Customer can analysis on a real time based usage of various electrical device and control consumption of energy through the app. The cost of this system is very less as compared to existing system in market. The feature that makes the application unique is that no other application has the facilities to measure power consumption of each device.

IV. SYSTEM ARCHITECTURE

The Current sensor which is a vital part of our system is attached to the phase wires of industrial units or home. Then it senses the current usage by industrial/home and give to arduino, where current sensors are attached to the arduino with help of extra circuit which contain 3.5mm audio connectors. This circuit helps to limit the voltage coming from industrial units/home because arduino have only 5v capacity. Current sensors give the analog value of current to arduino which convert to digital value within arduino (in build Analog to Digital Convertor). From this arduino values are passed to database through Wi-Fi module. Then data processes in database which provides data to the user through

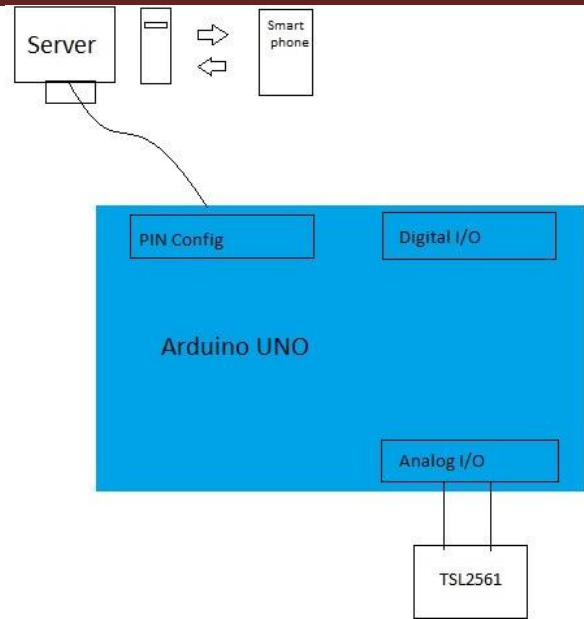


Fig. 3: Sensor Module outline

android application. The continuous monitoring values are given to user. Along with that the daily, monthly and yearly usage analyses are also given. The prediction of electricity bill given to user provides the energy consumption and cost estimation of industrial units [10].

1) *Power Analyzer Module:* The block diagram to show

the operation of the power analyzer is shown in figure 2. The current sensor SCT013030 is the main part of the circuit. The electricity measures are detected in real time and passed to the server through ESP 8266 WiFi module connected to the Arduino Nano. The SCT013 030 split core current sensor is capable of detecting a maximum current of 30A and provides a peak output voltage of 1V peak for that current. The output voltage thus generated is then transmitted to the Arduino Nano microcontroller via the input of the analog-to-digital converter (ADC). This voltage waveform is shifted up to 2.5 VDC. The rms value of the output signal is calculated and the power is calculated in the program.

2) *Home Automation Architecture:* Figure 4 depicts the

basic diagram of the proposed IoT Smart Home (IoT SHS) system. This is a low-cost, easily manageable, and profitable product. By providing

superior remote control for home appliances, it addresses the entire company’s segment.

To control devices remotely, it comes standard with WiFi. Lights, fans, and sockets are usually found in every room of any house where our items can be installed. This product does not affect the room’s electrical distribution wiring; everything remains the same except for the relays, which are wired in series with the switch or socket in the distribution box.

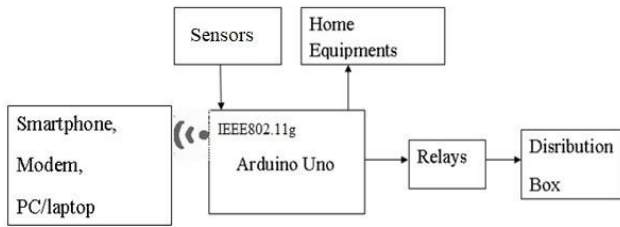


Fig. 4: Home Automation Architecture

The controller is a WiFi-based microcontroller (Arduino UNO) that serves as the system’s brain and controls all of the other components. The ambient temperature is indicated through a temperature sensor.

3) *Sensor Module:* The TSL2561 is a cheap, but sophisticated, light sensor. To better predict the human eye’s response, the TSL2561 integrates infrared and visible light sensors. This is contrast with simpler sensors, such as photoelectric sensors and photodiodes . The TSL2561 can measure both very small and very large amounts of light because it is a built-in sensor (it absorbs light for a pre calculated amount of time). The block diagram is shown in figure 3.

Circuit Diagram : The CT sensor cannot be directly connected to the Arduino since the high current from the

sensor can damage the Arduino, So we connect firstly the burden resistor of low resistance of about 33 is connected represented by R_b in the figure. Then the voltage divider bias consisting of R_1 and R_2 and additionally a bypass capacitor C_1 of 470F to block the digital signal, there by getting only the analog current signal to the Arduino. The analog pins of the arduino is indicated by notation $A_0 - A_5$ we have connected the CT sensor to pin A_0 . Then by clamping the CT sensor on alive wire gives the current output.

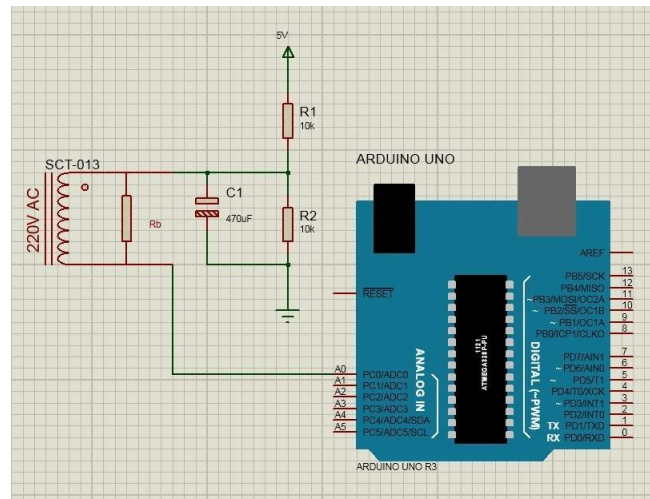


Fig. 5: (CT sensor interfaced with arduino

V. EXPERIMENTS AND DISCUSSION

Initially, login to the Digital space using username and password and check the username and password is valid by comparing in database. If it is valid it enter to the next module else unsuccessful login. There are different phases for the app where we can do the control of our home appliances automatically through the application.

Now we measure the usage statistics of the home that is read through the CT sensor. We can analyse the power usage in various manner according to our requirements. The app will provide the provision to calculate the power consumption in daily, monthly and year wise. Also have the facility to monitor the live power consumption. The app interface for the power analysis is shown in figure 6.

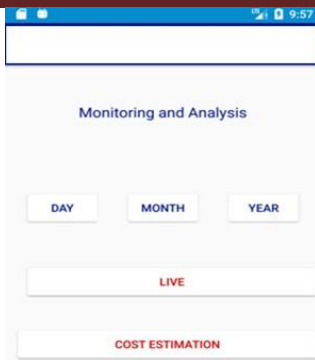


Fig. 6: Application Interface

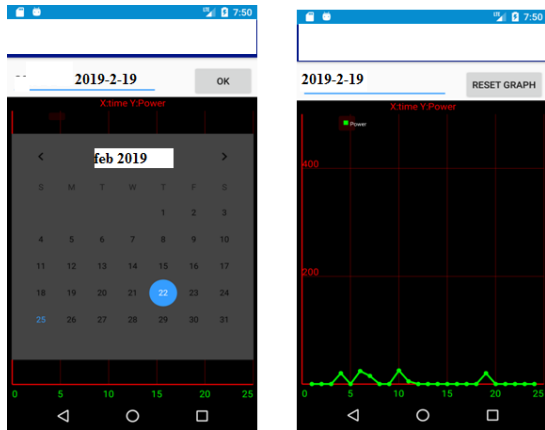


Fig. 7: Daily usage analysis

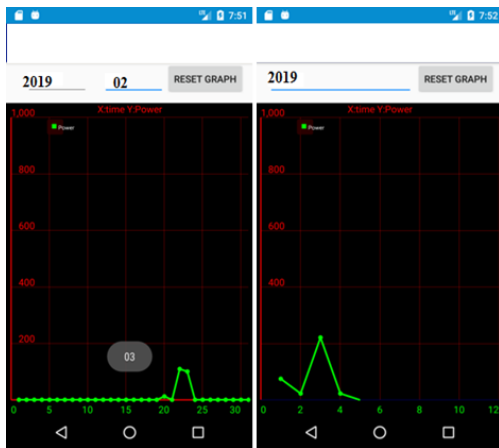


Fig. 8: Monthly and yearly usage analysis

For calculating the daily consumption we can choose the option in the app and it will give the consumption as a graph which is shown in the figure 7. By taking the power consumption in

this manner we would be able to control the usage.

We have also done the analysis of monthly and yearly power consumption rates and plotted as a graph. It is very convenient to track the miss usage of power. The results from the analysis is shown in the figure 8. The next part in the experiment was the cost estimation. It will give the summary of average power usage in a Day/Month/Year base. Also we have done the average cost estimation for a day, Month and Year.

CONCLUSION

IoT based energy the board framework can contribute a ton into preservation of energy. It can control different electric devices in home from anywhere at any time using internet connection or Wi-Fi. Similarly the cost estimation can be predicted for each industrial unit or home. We proposed an IoT Based Power Analyzer is a general purpose real time mobile application. It estimates the cost of each industrial unit or home. We propose creating a smart plug, which is an energy observation system that provides real-time information on energy use at the device level. An Arduino microcontroller board, an ENC28J60 LAN module, and a current electrical device sensing element are used in the suggested device. The end product is a smart plug that uses the Arduino-android platform to monitor a remote device. The second section focuses on the design of an IoT sensible House System (IoTSHS), which might provide access to a smart home via a mobile device, similar to a PC or laptop. The Arduino Uno microcontroller is the standard controller for the IoTSHS. By connecting sample appliances and successfully controlling them from a wireless mobile device, the home automation system has been experimentally proved to perform satisfactorily.

REFERENCES

[1] Aldabbagh, Ghadah, Raneen Alzafarani, and Ghadi Ahmad. "Energy Efficient IoT Home Monitoring and Automation System (EE-HMA)."
 [2] Kodali, Ravi Kishore, and Subbachary Yerroju. "Energy

- efficient home automation using IoT." In 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT), pp. 151-154. IEEE, 2018.
- [3] Jabbar, Waheb A., Tee Kok Kian, Roshahliza M. Ramli, Siti Nabila Zubir, Nurthaqifah SM Zamrizaman, Mohammed Balfaqih, Vladimir Shepelev, and Soltan Alharbi. "Design and fabrication of smart home with Internet of Things enabled automation system." IEEE Access 7 (2019): 144059-144074.
- [4] Gray, Chrispin, Robert Ayre, Kerry Hinton, and Leith Campbell. "'smart' is not free: Energy consumption of consumer home automation systems." IEEE Transactions on Consumer Electronics 66, no. 1 (2019):87-95.
- [5] Kitayama, Ryosuke, Takashi Takenaka, Masao Yanagisawa, and Nozomu Togawa. "Scalable and small-sized power analyzer design with signal-averaging noise reduction for low-power IoT devices." In 2016 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 978-981. IEEE, 2016.
- [6] Gao, Xiaofei, and Qin Zhou. "A low consumption DSP based power analyzer." In The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014), pp. 164-168. IEEE, 2014.
- [7] Hui, Li, Wang Gui-rong, Wei Jian-ping, and Duan Peiyong. "Monitoring platform of energy management system for smart community." In 2017 29th Chinese Control And Decision Conference (CCDC), pp.1832-1836. IEEE, 2017.
- [8] Chaudhari, Sneha, Purvang Rathod, Ashfaque Shaikh, Darshan Vora, and Jignasha Ahir. "Smart energy meter using Arduino and GSM." In 2017 International Conference on Trends in Electronics and Informatics(ICEI), pp. 598-601. IEEE, 2017.
- [9] Al-Ali, Abdul-Rahman, Imran A. Zualkernan, Mohammed Rashid, Ragini Gupta, and Mazin AliKarar. "A smart home energy management system using IoT and big data analytics approach." IEEE Transactions on Consumer Electronics 63, no. 4 (2017): 426-434.
- [10] Hlaing, Win, Somchai Thepphaeng, Varunyou Nontaboot, Natthanan Tangsunantham, Tanayoot Sangsuwan, and Chaiyod Pira. "Implementation of WiFi-based single phase smart meter for Internet of Things (IoT)." In 2017 International Electrical Engineering Congress (iEECON), pp. 1-4. IEEE, 2017.



EARN YOUR MBA

WWW.IIMPS.IN



Accreditation & Ranking



UGC / NCTE Approved.

INFO@IIMPS.IN

☎ 011-41005174

R
S
E
A
R
C
H
G
A
T
E
W
A
Y

STOP PLAGIARISM



Arogyam Ayurveda
Holistic Healing through herbs



A
R
O
G
Y
A
M
O
N
L
I
N
E

PARIVARTAN PSYCHOLOGY CENTER



COLOR PSYCHOLOGY : HOW COLOR AFFECT YOUR CHILD



- BLUE** Calms your Child's Mind & Body
- YELLOW** Promotes Concentration, Stimulates the Memory
- PINK** Evokes Empathy, makes your Child Calm
- RED** Excites and energizes your Child's body
- GREEN** Improves Reading speed and Comprehension

www.parivartan4u.com



Confuse about your children's future?

भारतीय भाषा, शिक्षा, साहित्य एवं शोध

ISSN 2321 – 9726

WWW.BHARTIYASHODH.COM



**INTERNATIONAL RESEARCH JOURNAL OF
MANAGEMENT SCIENCE & TECHNOLOGY**

ISSN – 2250 – 1959 (O) 2348 – 9367 (P)

WWW.IRJMST.COM



**INTERNATIONAL RESEARCH JOURNAL OF
COMMERCE, ARTS AND SCIENCE**

ISSN 2319 – 9202

WWW.CASIRJ.COM



**INTERNATIONAL RESEARCH JOURNAL OF
MANAGEMENT SOCIOLOGY & HUMANITIES**

ISSN 2277 – 9809 (O) 2348 - 9359 (P)

WWW.IRJMSSH.COM



**INTERNATIONAL RESEARCH JOURNAL OF SCIENCE
ENGINEERING AND TECHNOLOGY**

ISSN 2454-3195 (online)

WWW.RJSET.COM



**INTEGRATED RESEARCH JOURNAL OF
MANAGEMENT, SCIENCE AND INNOVATION**

ISSN 2582-5445

WWW.IRJMISI.COM



**JOURNAL OF LEGAL STUDIES, POLITICS
AND ECONOMICS RESEARCH**

WWW.JLPER.COM

JLPE



IMAGE SPLICING DETECTION - COMPARISON OF DMAC AND DMVN NETWORKS

Roshan Prasad
Student, Dept of CSE
SCMS School of Engg.
and Technology Ernakulam
roshanp60@gmail.com

Sarath J
Student, Dept of CSE
SCMS School of Engg.
and Technology Ernakulam
jsarathkaralmanna@gmail.com

Sten Benny
Student, Dept of CSE
SCMS School of Engg.
and Technology Ernakulam
stenbenny02@gmail.com

Vinay Stephen
Student, Dept of CSE
SCMS School of Engg.
and Technology Ernakulam
vinaystephen1@gmail.com

Litty Koshy
Asst Professor, Dept of CSE
SCMS School of Engg. and Technology, Ernakulam
littykoshy@scmsgroup.org

Abstract---Constrained image splicing detection and localization (CISDL) is a difficult task for image forensics that examines two input suspicious pictures and determines whether one contains suspected portions copied from the other. Here a unique adverse learning approach for training the deep matching network for CISDL is presented. The goal of the deep matching network based on atrous convolution (DMAC) is to create two high-quality candidate masks that show the suspicious regions of the two input pictures. The correlation layer based on the skip architecture is proposed to capture hierarchical features in DMAC, and Atrous spatial pyramid pooling is used to extract features with rich spatial information. Another model called DMVN uses the same process as DMAC but it is not use atrous convolution. A comparative study of both models was done ,in which the DMAC model is better because it gives high resolution fined grained mask.

Keywords:Atrousconvulution,DMAC,DMVN,CISDL

I. INTRODUCTION

Malicious image forgery is becoming a global epidemic in recent years, due to the rapidly declining cost of digital cameras and quick development of sophisticated image editing tools. Forgers may use forged images to produce fake news, spread rumors or give false testimony, which result in negative social impacts. Image forensics, which seeks to distinguish forged images and prevent forgers from using forged images for unscrupulous business or political purposes, has attracted great attention in research and industrial communities. A variety of image forensics

methods investigate an individual image and detect its high-level or low-level inconsistencies caused by image manipulation. However, it is still a challenging task to accurately distinguish forged images, due to advanced image manipulation techniques and limited information provided by a single image. Moreover, these image forensics methods identify forged images or regions without providing the source of forged regions or specific tampering process, but these auxiliary evidences can provide more clues and make results more convincing in real applications. Constrained image splicing detection and localization (CISDL) is newly formulated in the Media Forensics Challenge. Different from “conventional” splicing detection, “constrained” means that the inputs are two images: one is a probe image and the other is a potential donor image. In CISDL given a probe image P and a potential donor image D, CISDL aims to detect if a region of D has been spliced into P, and consequently provide mask images P_m and D_m indicating the regions of P were spliced from D. DMVN generates correlation maps by comparing high-level low-resolution feature maps of VGG, and constructs an inception-based mask convolution module to locate suspected regions. However, low-resolution feature maps restrict DMVN's ability to detect accurate boundaries and small suspected regions. Here proposed a deep matching

network based on atrous convolution (DMAC) to generate high-quality candidate masks from high-resolution feature maps.

The basic DMAC architecture achieves significant improvements over DMVN.

This work, proposes a DMAC network which takes two images as inputs. These input images are fed into a atrous convolution network for feature map extraction. The extracted feature maps fed into the correlation layer and atrous spatial pyramid pooling for feature maps comparisons.



Fig 1. a) Donor Image b) Donor Mask c) Probe Image d) Probe mask

II. LITERATURE SURVEY

In [1] Proposes a new convolutional layer that suppresses image content and learns forgery detection. In [1] they proposed a CNN to learn manipulation detection features directly from data and it is used in image forensics. In [2] DMAC is combined with adversarial learning for effective image forgery detection. In [3] propose an optimized 3D lighting estimation method by incorporating a more general surface reflection model. In [4] propose a framework to improve the performance of forgery localization via integrating tampering possibility maps. In [5] proposed algorithm automatically computes a

likelihood map indicating the probability for each 8×8 discrete cosine transform block of being doubly compressed.

III. PROPOSED METHODOLOGY

In this section, explains the proposed framework, as shown in fig 2. In the DMAC model there are three modules namely feature extraction module, correlation module and ASPP (Atrous Spatial Pyramid Pooling). In the feature extraction module atrous convolution is adopted to enrich the spatial information of convolutional features.

In the correlation module, the skip architecture is designed for hierarchical features comparisons and in the ASPP module is used to capture the information of different scales, Atrous Spatial Pyramid Pooling is constructed to generate the final mask. ASPP contains multiple parallel atrous convolutional layers with different sampling rates.

In DMAC Model, two images as inputs probe and donor. In which donor is the original image which is captured by camera also called as authentic image and probe image is the image containing spliced portion of donor image also called as tampered image. This two image is given as input to the model and it produce high resolution fine grained mask as output. The DMAC model is using atrous convolution which is used to give high resolution mask. At last creating another model called DMVN, in this which is not using atrous convolution and compare with DMAC model for accuracy.

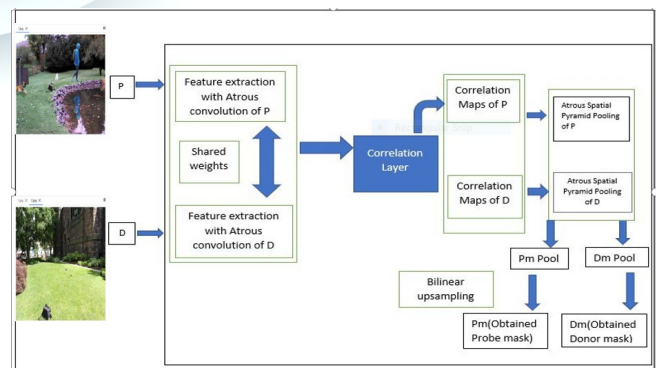


Fig 2. Proposed System Architecture

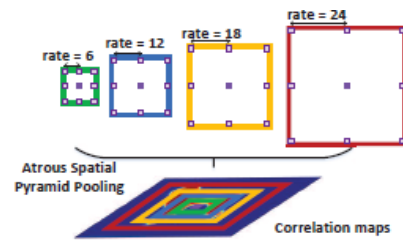


The DMAC network is a unique adversarial network in which feature extraction modules employing atrous convolution, the correlation layer with skip architecture, and ASPP are designed to enrich geographical information. In DMAC atrous convolution, the correlation layer and ASPP are used to capture hierarchical properties and localise impacted regions at many scales, respectively. The detection network and discriminative network, which act as losses with supplementary parameters, monitor DMAC's adversarial training.

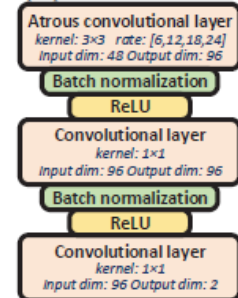
A. Feature Extraction with Atrous Convolution

CNNs' pooling or downsampling techniques necessarily degrade the spatial resolution of the output feature maps. As a result, in this research, atrous convolution is employed to create high-resolution feature maps. Atrous convolution allows us to vary the field-of-view of filters by adjusting the rate value without adding any more parameters. This module alters the image by adjusting the colour, contrast, and light intensity, which aids in the creation of a high-resolution mask.

Assume the input feature maps are scaled down by a factor of two before being convolved with standard convolution filters. The created feature maps are just a fourth the size of the original feature maps, and traditional filters only acquire answers from a quarter of the image locations. If we eliminate the downsampling layer and directly convolve the input feature maps, the filters will have a smaller field of vision. Fortunately, we may keep the original field-of-view by employing atrous convolution with rate $r = 2$. Using atrous convolution techniques, we can create high-resolution feature maps, get all answers from the input feature maps, and don't need any additional parameters or calculation. Despite the fact that the effective filter size increases, we only need to examine non-zero filter values, resulting in a constant number of filter parameters and operations per site.



(a) Multiple parallel atrous convolutional layers.



(b) The ASPP separate branch.

B. Correlation Computation Module

To build dense high-resolution feature maps, atrous convolution is used as the basic process. One of the most difficult difficulties in deep matching tasks is deep feature comparison. For other tasks, only the neighbouring fields are compared, allowing for the creation of complex correlation layers. They usually compute the scalar product of a pair of individual descriptors at each place for long-range correlation computing tasks. We can denote correlation computation procedure as a function. The skip architecture is proposed to effectively organize the atrous convolution and hierarchical convolution features. It makes full use of the feature extraction module's wealth of information. Three sets of feature maps are produced by the atrous convolution layer: f_3 , f_4 , and f_5 . As a result, three sets of correlation feature maps can be created using the feature maps f_3 , f_4 , and f_5 , with no upsampling or mapping functions required. The computation procedure of the proposed correlation layer based on the skip architecture can be summarized as Algorithm: The skip architecture is used to compute the correlation layer.

C. Atrous Spatial Pyramid Pooling

The fact that the tampered portions are on different scales is a problem in the image splicing detection technique. The final masks are generated using Atrous Spatial Pyramid Pooling (ASPP), which captures the information



on multiple scales provided by the correlation maps. Simply ASPP is a discriminative network that drives the DMAC network to produce masks that are hard to distinguish from ground truth ones. Multiple atrous convolutional layers with varying sampling rates are present in ASPP. As a result, those obscene convolution filters have varying field-of-views and can focus on altered parts of various scales. A separate branch of convolutional layers, batch normalisation, and ReLU layers follow each atrous convolutional layer with one sample rate. The individual branches are then merged to create the finished masks. During mask formation, there is no upsampling operation with learnable parameters, thus we just use bilinear upsampling during test time.

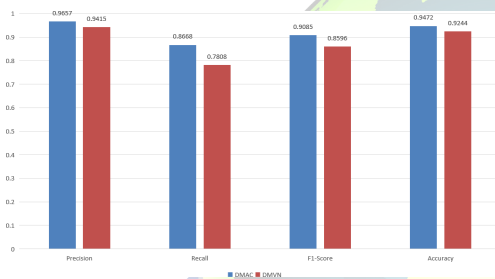
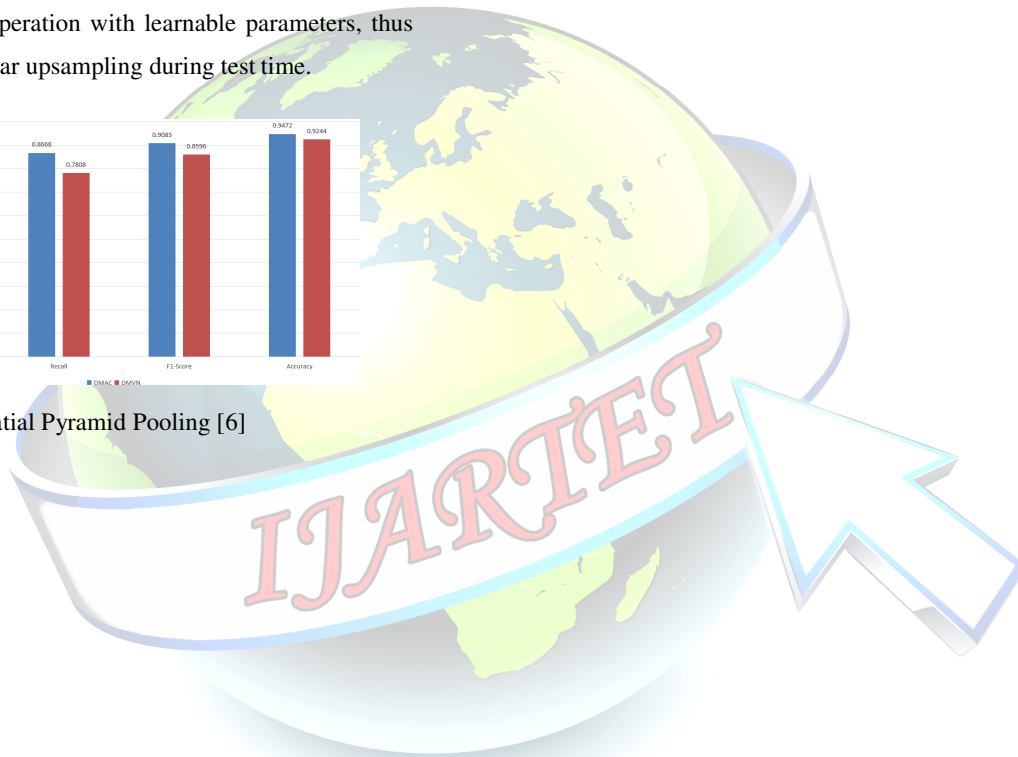


Fig 3: Atrous Spatial Pyramid Pooling [6]





IV. RESULT AND DISCUSSION

Precision: 0.8541114489666498
Recall: 0.9286295170072838
F1 Score: 0.8898130659054911
Accuracy: 0.9384002685546875

Fig 4- Result of DMAC Model

Fig 5- Comparison Between DMAC and DMVN

V. CONCLUSION

This work provides a simple but effective framework for detecting image splices. A unique adversarial learning framework is proposed to deal with the CISDL task. To improve the DMAC network's ability to detect small matching regions and multi-scale regions, atrous convolution, the skip architecture, and ASPP are used. A lot of experiments are conducted on all generated datasets and also all publicly available datasets. The experimental results demonstrate the appealing performance of the proposed adversarial learning framework and the DMAC network. The use of atrous convolution and ASPP has clearly increased the effectiveness of the algorithm compared to the existing ones. Although the techniques to detect small tampered regions and regions under huge changes still need further research.

REFERENCES

- [1] Belal Ahmed, T. Aaron Gulliver, Saif alZahir (2020) "image splicing detection using mask-RCNN", "Verlag London Ltd., part of Springer Nature 2020".
- [2] Snigdha K. Mankar, Prof. Dr. Ajay A. Gurjar (2015) image splicing detection based on SVM (Support Vector Machines) classifier at IETE Amravati Center Volume 5.
- [3] A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, and others, "Splicing image forgery detection based on DCT and Local Binary Pattern," in Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE, pp. 253–256, 2013.
- [4] Hsu Y-F, Chang S-F. (2007). Image splicing detection using

camera response function consistency and automatic segmentation. In: Multimedia and Expo, IEEE International Conference on, p. 28–31.

[5] Chi-Man Pun, Bo Liu, Xiao-Chen Yuan, "Multi scale noise estimation for image splicing forgery detection" in Journal of visual communication and image representation, pp. 195–206, 2016.

[6] Y. Liu, X. Zhu, X. Zhao and Y. Cao, "Adversarial Learning for Constrained Image Splicing Detection and Localization Based on Atrous Convolution," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 10, pp. 2551–2566, Oct. 2019, doi: 10.1109/TIFS.2019.2902826.

[Home](#) > [Cluster Computing](#) > Article

[Published: 29 September 2021](#)

Evaluation of adversarial machine learning tools for securing AI systems

[S. Asha](#) & [P. Vinod](#) 

Cluster Computing, **25**, 503–522 (2022)

834 Accesses | **1** Citations | [Metrics](#)

Abstract

Artificial intelligence aims to build intelligent systems capable of performing tasks that need human intelligence. Research works in recent years have revealed many potential vulnerabilities in machine learning algorithms. Precisely to exploit these vulnerabilities, an attacker may attempt to design an adversarial input to be incorrectly processed by machine learning algorithms. This paper focuses on methods of generating adversarial samples and discusses possible countermeasures. Our proposed method effectively checks the robustness of different machine learning models using six available hostile robustness tools and summarizes their current development state. The work compares the features

Author information

Authors and Affiliations

Department of Computer Science & Engineering, SCMS School of Engineering & Technology (Affiliated to APJ Abdul Kalam Technological University, Trivandrum, India), Ernakulam, India

S. Asha

**Department of Computer Applications,
Cochin University of Science & Technology,
Cochin, India**

P. Vinod

Corresponding author

Correspondence to [P. Vinod](#).

Additional information

Publisher's Note

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

Appendix

List of notations used in the paper

15 November
2020

22 July 2021

15 September
2021

Published

29 September
2021

Issue Date

February 2022

DOI

<https://doi.org/10.1007/s10586-021-03421-1>

Keywords

Adversarial machine learning **Evasion attack**

Poisoning attack **Adversarial robustness tools**

Machine learning models



IMPACT ANALYSIS OF THE COVID19 ON THE ATMOSPHERIC AIR QUALITY AND ELECTRICITY CONSUMPTION PER DAY IN INDIA.

Dr. Sreelekha B.

Associate Professor, SCMS School of Engineering and Technology, Angamaly, Kerala, India.

Nandini Menon

Student, Vellore Institute of Technology, Vellore, Tamil Nadu, India.

ABSTRACT The corona Virus pandemic is an unexpected part of our life. Following the lockdown procedures implemented by the Indian government, the atmospheric air quality and the power sector of India noticed decadent changes. However, the relaxation in lockdown brought situations almost back to normal. The paper focuses on analyzing these changes to review the current scenario according to the published studies. It focuses on electricity consumption, night light intensity, and variations in many pollutants like NO₂ and PM. For ease, the main cities chosen for the research are the metropolitan cities of India. The data are from US-EPA, POSOCO, VIIR satellites and so on. The paper helps us understand if this variation is a boon or bairn for the Indian economy.

KEYWORDS : COVID-19, Electricity Consumption, Air Quality, Atmosphere, Energy Consumptions, Night Light Intensity.

INTRODUCTION

On March 25th, Prime Minister of India Narendra Modi announced a nationwide lockdown, accounting for the safety precautions required to fight against the Coronavirus (SARS-CoV-2). Ever since then, there has been an exponential surge in the number of infections reported per day. From April to September 2020, as the number of cases kept increasing, more people restricted themselves from going out to keep themselves safe and well. As of September 7th, 2020, India recorded 41.13 Lakhs of confirmed infections, making her the second most affected country in the world after the USA. On July 15th, 2020, the phase-1 clinical trials for the first indigenous Coronavirus Vaccine, Covaxin, developed by Hyderabad-based pharmaceutical company Bharat Biotech and the National Institute of Virology and Indian Council of Medical Research, starts across the country.



Fig: 1 (6)-Corona-Virus Pandemic India-Timeline.

Simultaneously, many researchers have been working on the impact of lockdown on atmospheric CO₂ levels and electricity consumption per day. While e-collaborations positively impact the present climate and environment, it is definite that reduction will be short-lasting, attributing it to the close-down of transport, construction works, and industrial activities.

The Corona-Virus pandemic has an unparalleled effect on our everyday life, which will continue until a minimum of the next three years. While research, vaccination, and protocol documentation procedures are currently ongoing, the impact of lockdown on the environment is also a widely inspected topic. A calculated set of restrictions imposed on the economy to reduce the spread of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS- CoV-2) has an overall positive effect on the environment. The beneficial impact includes reduced particulate matter levels in the atmosphere, decreased carbon dioxide (CO₂) levels, reduced kerosene or related fuel use, and increased awareness about the importance of the 5Rs. A study conducted by analyzing the data and images collected from the Sentinel satellite-ESA revealed a 45% decline in atmospheric NO₂ levels in India(1). Air pollution takes the lives of almost 1.7 million per year. Besides global warming, Air pollution has fueled many recent disasters, namely the Kerala floods of 2018-2019, the Assam floods 2019-2020, the Amazon Forest Fire, the Sydney forest fires, Australia floods 2021, and it keeps going on. Air pollution symptoms include aggravated respiratory diseases like asthma and bronchitis, dry throat, wheezing, nausea, and headache. In India, the Northern parts are the most polluted areas, especially Delhi, mainly due to emissions from Vehicles, brick kilns, coal-based thermal power plants, and crop remnants(2). The total energy consumption (ameasure for the amount of electricity consumed) and

the lights per area (a measure of the intensity of light in the area) are the proxy indicators for consumption level measurements (3). Ever since the nationwide lockdown, there is a decline in daily energy consumption. The official power consumption data captured by POSOCO (Power System Operation Corporation) has recorded a 26% decline since the nationwide lockdown. The table below represents a section of the data released by POSOCO (4).

Date	Energy Consumption (GWh)					All India
	Northern Region	Western Region	Southern Region	Eastern Region	North-Eastern Region	
18-Mar-20	827	1187	1148	383	42	3586
22-Mar-20	734 (-11%)	971 (-18%)	975 (-15%)	315 (-18%)	36 (-13%)	3030 (-15%)
23-Mar-20	724 (-12%)	996 (-16%)	1030 (-10%)	325 (-15%)	39 (-5%)	3113 (-13%)
24-Mar-20	695 (-16%)	944 (-20%)	983 (-14%)	314 (-18%)	39 (-7%)	2975 (-17%)
25-Mar-20	665 (-20%)	844 (-29%)	911 (-21%)	320 (-16%)	36 (-13%)	2777 (-23%)
26-Mar-20	628 (-26%)	771 (-34%)	891 (-23%)	327 (-15%)	35 (-16%)	2652 (-26%)

Figures in parentheses indicate percentage change from 18th March 2020

Table 1.1 (4)-POSOCO Data.

The data collected from the different sources suggests a pollution reduction, but the trend is inconsistent. Analyzing these trends and converting them into more consistent data is of the highest priority right now. COVID19 is posing as an opportunity to do the same. The paper has put together a review of the recent researches analyzing this impact.

ATMOSPHERIC AIR QUALITY

Researching and identifying air quality index measures of a particular area requires humongous data on the primary sources in the area, optimal pollution levels, meteorology, demographics, geography, and computational capacity (7). PM-2.5 and PM-10 are the most dangerous types of pollutants released into the atmosphere. PM-2.5 is a type of particulate matter whose size is less than 2.5 micro-meter, which is small enough to enter our lungs and bloodstream. Numerous studies link PM-2.5 to various health risks compared to other pollutants.

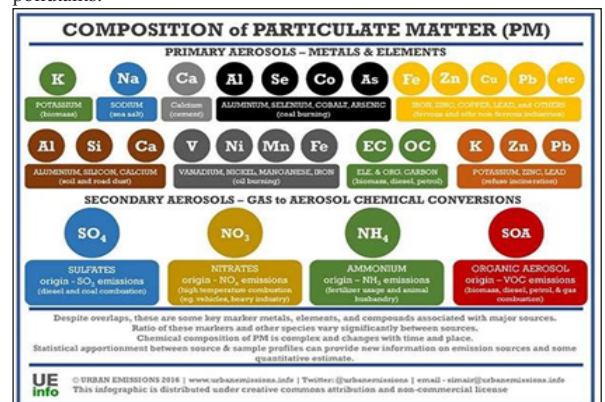


Fig: 2 (7)-Composition of Particulate Matter.

Table-2 (8)- Percentage Reduction In Emissions During The Lockdown Period 25/03/2020 To 15/04/2020 Compared To 24/03 To 15/04, 2019.

Pollutant	Date	Bengaluru	Chennai	Hyderabad	Mumbai
NO ₂	24, Mar–15 Apr,2019	67.1	47.6	39.7	6.8
	BaU 25, Mar–15 Apr,2020	57.7	28.7	30.1	27.9
PM _{2.5}	24, Mar–15 Apr,2019	45.1	45.7	18.9	42.0
	BaU 25, Mar–15 Apr,2020	45.2	28.7	12.3	39.9
SO ₂	24, Mar–15 Apr,2019	1.7	8.2	9.4	3.4
	BaU 25, Mar–15 Apr,2020	10.7	33.0	-17.2	45.2
CO	24, Mar–15 Apr,2019	23.2	39.6	24.6	-55.1
	BaU 25, Mar–15 Apr,2020	27.6	13.4	9.8	37.1

CHANGE IN PM-2.5 AND PM-10

The inevitable parameter in determining the air quality of a particular area concerning PM is identifying the optimal ambient concentration levels of the pollutant. Table 2 shows the percentage reduction in emissions during the lockdown period 25/03/2020 to 15/04/2020 compared to 24/03 to 15/04, 2019. (Eregowda, n.d.)(8) tabulated the results by collecting data from Bengaluru, Chennai, Hyderabad, and Mumbai. Table-2 shows a 5.1, 45.7, 18.9, and 42% decrease in PM-2.5 pollutant emission. (Ghosh, n.d.) (9) has tabulated the results by collecting data from the Indian Metropolitan cities NCR-Delhi, Mumbai, Chennai, and Kolkata. Figure 3 shows the results obtained by Ghosh, n.d. from the Landsat 8 OLI and TIRS- Derived Data and Mamdani Fuzzy Logic Modelling Approach to understand PM-10 concentration Variation.

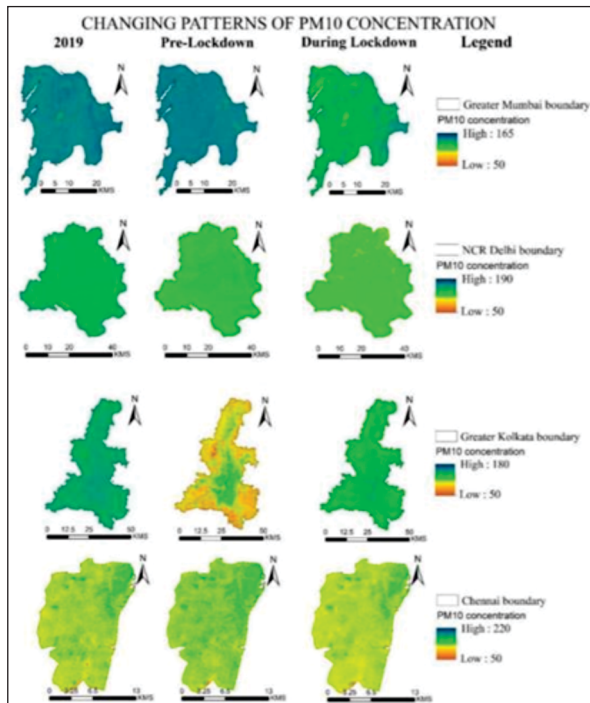


Fig: 3 (9)- Changing PM-10 concentration comparison between 2019, pre-lockdown 2020, and during lockdown 2020.

The PM-10 sources include motor vehicles and construction works. It is the causes numerous health risks, environmental harm, and reduced human comfort levels. While the concentration legends for Mumbai and Chennai are decreasing, Delhi and Kolkata show an increase in PM-10 concentration between pre-lockdown and during the lockdown (9). The variation may be due to the shutdown of industries and

restricted human movement compared to other cities. Mr. P. Singh, in his research work, focused on the Air Quality Index and PM2.5 levels by collecting data from the cities with a branch of the US Embassy in India (10). The Embassies collect the data via US Environmental Protection Agency (EPA) through the Air-Now portal (11).

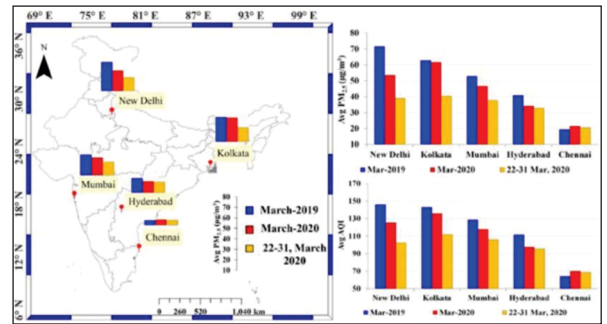


Fig: 4 (10)- The left panel shows the locations for the research work done by Mr. P. Singh. The right panel presents the Average PM2.5 and Air Quality Index (AQI) chart for the area.

Analyzing the data represented in fig-1, it's clear that the PM2.5 and AQI at the areas of study as reduced noticeably during the lockdown period. The pollution levels at Mumbai, Hyderabad, and Chennai have decreased by 19.25%, 3.99%, and 5.40%, respectively. At the same time, Kolkata and Delhi show a considerable reduction of 34.52% and 27.57%, respectively, in the pollutant levels. The northern parts of India, especially the Indo-Gangetic Plains (IGP), have higher levels of PM2.5 throughout the year. The factors included demographic, geographic, seasonal activities, and meteorological parameters. The proximity to the sea for Chennai and Mumbai can cause air mass circulation from the sea surface, which is a possible explanation for the reduced pollution levels. The same is applicable for New Delhi and Kolkata.

CHANGE IN NO2

NO2 is one of the most common pollutants which is highly dependent on the local sources. The dependency is due to its short residence time in the atmosphere. Ms. Eregowda stated in her paper that the NO2 concentration levels at Bengaluru fell from 50 µg/m3 to 10 µg/m3 throughout the lockdown period. The same follows for Chennai, Hyderabad, and Pune. Figure 5 presents the above data graphically.

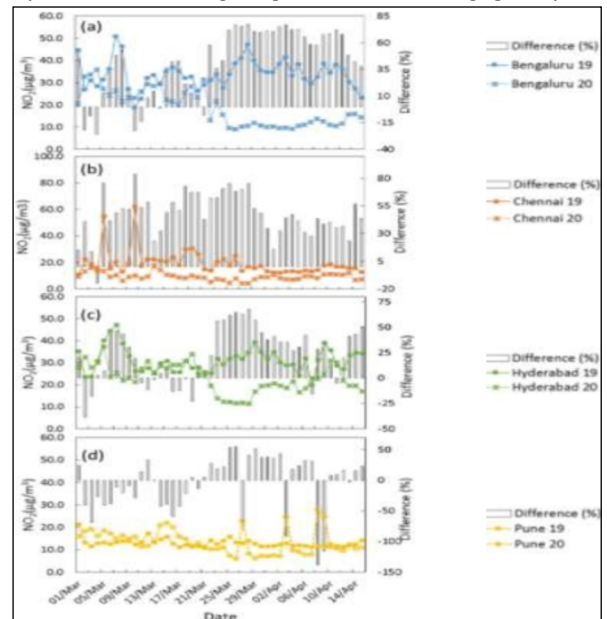


Fig-5 (8)- NO2 Level Graphs With % Difference Representation.

Mr. P. Singh considered the tropospheric NO2 measure by analyzing the data from the Ozone Monitoring Instrument (OMI). The Ozone Monitoring Instrument (OMI) is a section of NASA's A-Train Satellite that measures the levels of various atmospheric gas concentrations. Table-3 represents the box coordinates of the US-embassy locations chosen by Mr. Singh for the research.

Table-3 (10)- The Locations Of The Us-embassies And Their Box Coordinates From OMI.

Location	Latitude	Longitude	Box Coordinates
Delhi	28.59	77.18	W-76.68, S-28.07, E-77.68, N-29.07
Kolkata	22.54	88.35	W-87.86, S-22.08, E-88.86, N-23.08
Mumbai	19.06	72.86	W-72.42, S-18.55, E-73.42, N-19.55
Hyderabad	17.44	78.47	W-77.78, S-17.01, E-78.78, N-18.01
Chennai	13.05	80.25	W-79.76, S-12.56, E-80.76, N-13.56

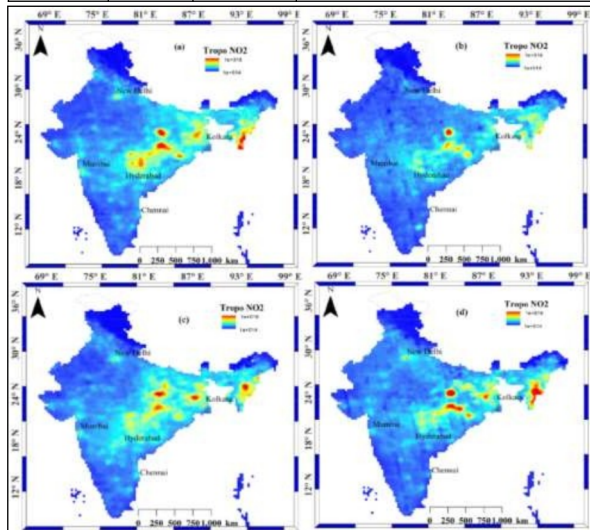


Fig- 6- Tropospheric NO₂ (spatial Variations)- a)10–21March 2019, (b)10–21 March 2020, (c)22–31 March 2019, and (d)22–31 March 2020.

Fig-6 represents the results tabulated by Mr. P. Singh. An HYPSPPLIT Model prepared by Mr. P. Singh shows that long-range air mass transportation affects the air quality at the five selected locations. The westerly air mass transfer is what affects Kolkata, a city located in the eastern Indo-Gangetic Plain. The sources of NO₂ in Mumbai, Delhi, and Kolkata are anthropogenic. While so, the release of NO₂ in Chennai and Hyderabad is due to the burning of biomass. Hence the decline in NO₂ levels during the lockdown is mainly due to the reduction in anthropogenic emissions.

CHANGE IN LAND SURFACE TEMPERATURE(LST)

LST is an important factor for environmental health, as it depends on numerous physical and atmospheric parameters (9). Factors like cloud conditions, month, Land Use/Land Cover (LULC) patterns, and so on governs the results. The LST map for 2019, before lockdown 2020 and after lockdown 2020, is given in Figure-7. The transition period in India from winter to summer is from February to March. And the summer season is from March to May.

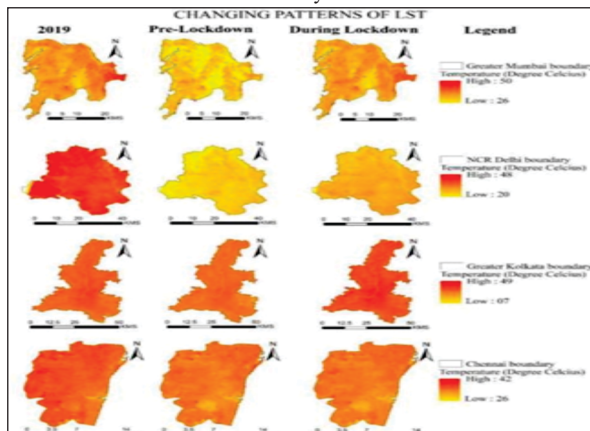


Fig- 7-(9) Changing variations in LST. The LST Map.

These variations in seasonal temperatures can show an increasing trend in the land surface temperature. From the map, it is clear that the temperature map during the pre-lockdown period is less compared to the others. Even though the maps show a similar trend in the four cities, the temperatures in Mumbai, Delhi, and Kolkata are around 48°C to 50°C. Chennai records lower temperatures in the range of 42°C to 44°C.

ELECTRICITY CONSUMPTION

India is the second most populated (1,391.99 Million) and the seventh-largest country in the world(16). The lockdown has imposed several restrictions on many industrial and everyday goods and other service-related activities. Between January and February of 2019 and 2020, the energy demand increased by 3% and 7%, respectively (17). But, during March 2020, the supply-demand reduced by 3%. Between March 24th and April 19th, the power supply decreased by 25%.

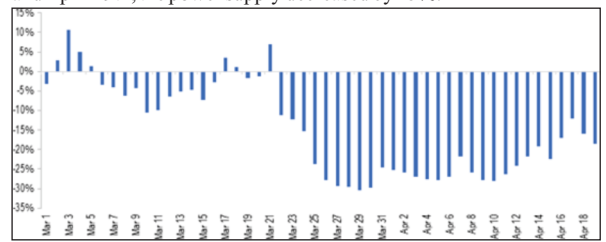


Fig- 8 (17)-% change in power supply between March 1st to April 19th.

It is necessary to understand the power consumption in different sectors of India to analyze the power demand variation in India. Fig-9 presents the power consumption in consumer segments of India in 2018-19. Since the government forced the industrial and commercial sectors to shut down during the lockdown, the decrease in demand is self-explained. Table- 4 shows the contribution of different sources to energy production in India.

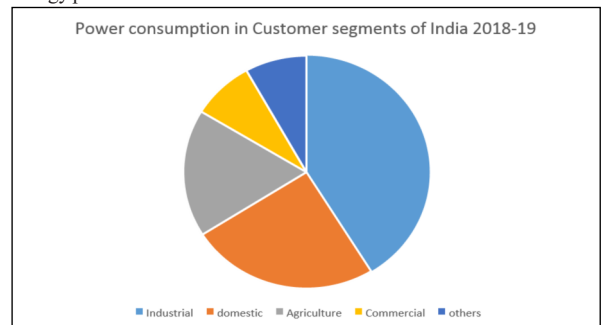


Fig- 9 (17)- Power consumption in Customer segments of India 2018-19.

Table-4 (17)- The contribution of different sources to energy production in India from March 1st to April 19th, 2021.

Energy Sources	Average Generation			Contribution to Total (in %)	
	Mar 1-Mar 24	Mar 25-Apr 19	% change	Mar 1-Mar 24	Mar 25-Apr 19
Coal	2,511	1,873	-25%	72.5%	65.6%
Hydro	302	331	10%	8.7%	11.6%
Renewables (of which)	325	312	-4%	9.4%	10.9%
a. Solar	157	162	3%	4.5%	5.7%
b. Wind	97	96	0%	2.8%	3.4%
Gas, Naptha, Diesel	132	146	11%	3.8%	5.1%
Nuclear	113	114	1%	3.3%	4.0%
Lignite	82	78	-5%	2.4%	2.7%
Total	3,465	2,854	-18%	-	-

Amid the lockdown, the power generation reduced by 25%, compensating for the decrease in demand (17). While considering electricity variations, night light intensity is also a contributing factor. Night light intensity provides information regarding energy consumption in areas with high spatial granularity. Electricity consumption and night light intensity are contributing factors for the analysis of national GDP (18).

For instance, the demonetization in India, 2016 was highly backed up by night light intensity analysis (19). As mentioned, the impact of the nationwide lockdown on India remained even after the release of a few restrictions. The consumption levels were below 14%, and the average monthly fluctuations remain 6% to 10% below the normal (18). Fig-10 shows the trend in electricity consumption from 2013 to 2020.



Fig-10 (18) The trend in electricity consumption in India from 2013 to 2020

C. M. Beyer extracted the nighttime light data from the VIIRS-DNB Cloud made available by the Earth Observation Group at the National Oceanic and Atmospheric Administration (NOAA). The data collected was from April 2013 to April 2020 (18). VIIRS satellites have a resolution of 15-arc seconds. Fig-11 shows the changes in the night light time trends in India.

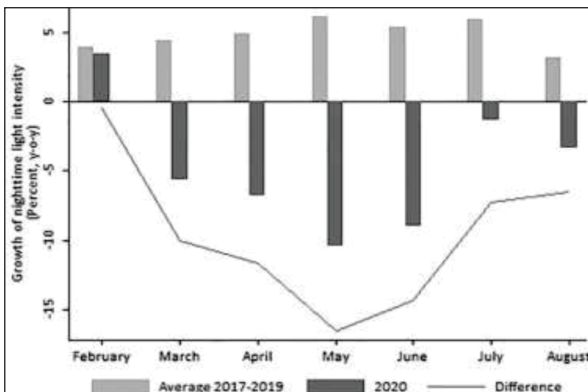


Fig-11 (18) the Changes In The Night Light Time Trends In India.

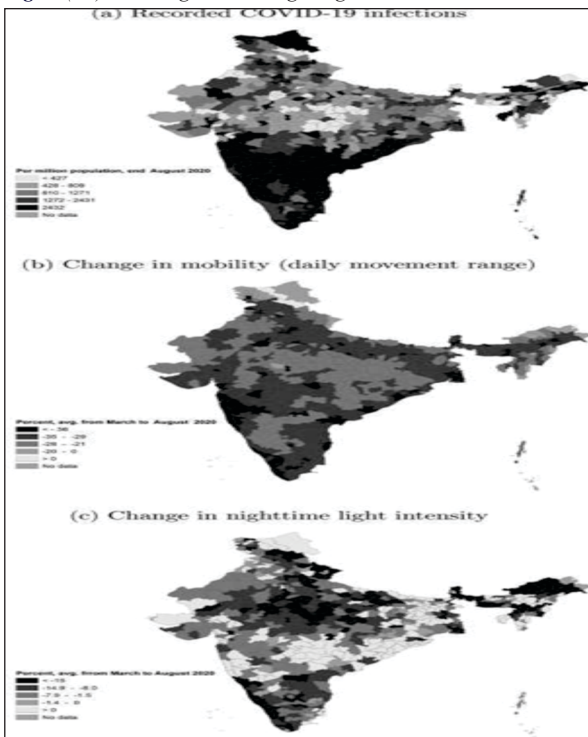


Fig-12 (18)- The spatial variation and impact of COVID-19 across India.

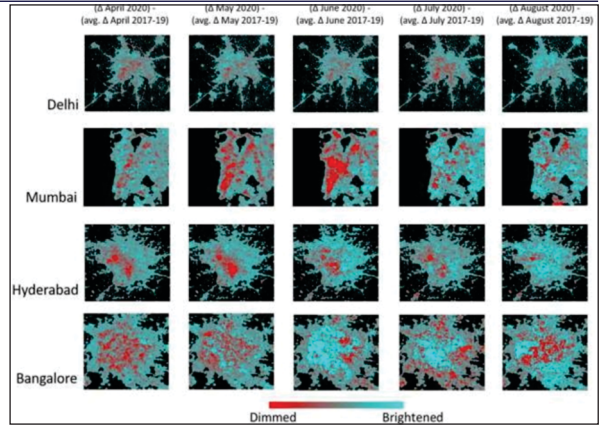


Fig- 13-(18)- The changes in night light intensity across India during the lockdown.

CONCLUSION

The impact of lockdown on India is a huge game-changer. The decrease in pollutant levels and electricity consumption rates are posing as an opportunity and a threat at the same time. Through this paper, we have discussed the findings achieved by the researchers for the betterment of our environment. India can rewrite its future to become a sustainable country. The documented results are proof highlighting our potentials.

As the COVID virus spread out, the government keeps extending the lockdown. The positive impact through this episode is huge. Since the energy consumption levels are proportional to household income, the reduced consumption can show the deteriorating levels of the economy. The Corona Virus pandemic will negatively impact all industries, including the power sector of India. But, so far, the impact is positive. Now we get to decide if it will remain the same or not.


REFERENCES

- (1) Shehzad et al., 2020 K. Shehzad, M. Sarfraz, S.G. Meran Shah The impact of COVID-19 as a necessary evil on air pollution in India during the lockdown Environ. Pollut., 115080 (2020), 10.1016/j.envpol.2020.115080 Google Scholar.
- (2) Venkataraman C, Brauer M, Tibrewal K, Sadavarte P, Ma Q, Cohen A, Chaliyakunnel S, Frostad J, Klimont Z, Martin RV, Millet DB (2018) Source influence on emission pathways and ambient PM2.5 pollution over India (2015–2050). Atmos Chem Phys Discuss 8:8017–8039 <https://acp.copernicus.org/articles/18/8017/2018/Article> Google Scholar.
- (3) <https://blogs.worldbank.org/endpovertyinsouthasia/indias-electricity-consumption-data-shows-economic-impact-covid-19>
- (4) <https://energy.economicstimes.indiatimes.com/news/power/coronavirus-impact-within-ten-days-26-per-cent-fall-in-indias-energy-consumption/74854825>
- (5) Positive effects of COVID-19 lockdown on air quality of industrial cities (Ankleshwar and Vapi) of Western India Ritwik Nigam, Kanvi Pandya, Alvarinho J. Luis, Raja Sengupta & Mahender Kothe. <https://www.nature.com/articles/s41598-021-83393-9>
- (6) Air pollution in Indian cities: Understanding the causes and the knowledge gaps Q&A with Dr. Sarath Guttikunda-14th December 2017, <https://www.cprindia.org/news/air-pollution-indian-cities-understanding-causes-and-knowledge-gaps>
- (7) Impact of lockdown associated with COVID19 on air quality and emissions from transportation sector: A case study in selected Indian metropolitan cities. Tejaswini Eregowda, Pritha Chatterjee & Digvijay S. Pawar, *Environment Systems and Decisions* (2021).
- (8) Impact of COVID-19 Induced Lockdown on Environmental Quality in Four Indian Megacities Using Landsat 8 OLI and TIRS-Derived Data and Mamdani Fuzzy Logic Modelling Approach. Sasanka Ghosh, Arijit Das, Tusar Kanti Hembram, Sunil Saha, Biswajeet Pradhan, andAbdullahM.Alamri. https://www.researchgate.net/publication/342749510_Impact_of_Lockdown_on_Air_Quality_in_India_During_COVID-19_Pandemic
- (9) Impact of Lockdown on Air Quality in India During COVID-19 Pandemic-July 2020Air Quality Atmosphere & Health 13(13) DOI:10.1007/s11869-020-00863-1, Project: Air Quality measurement during COVID-19, https://www.researchgate.net/publication/342749510_Impact_of_Lockdown_on_Air_Quality_in_India_During_COVID-19_Pandemic
- (10) <https://www.airnow.gov/>
- (11) Air Pollution in India: Real-time Air Quality Index Visual Map. <https://aqicn.org/map/india/>
- (12) AirNow for air quality data. <https://www.airnow.gov/>
- (13) OMI-NASA's A-Train Satellite. <https://giovanni.gsfc.nasa.gov/giovanni/>
- (14) INDIA: AIR QUALITY STANDARDS. <https://www.transportpolicy.net/standard/india-air-quality-standards/>
- (15) India-Wikipedia. <https://en.wikipedia.org/wiki/India>
- (16) Impact of COVID-19 on the Power Sector, Saket Surya. (The PRS Blog-April 23, 2020) <https://www.prsindia.org/theprsblog/impact-covid-19-power-sector>
- (17) Examining the economic impact of COVID-19 in India through daily electricity consumption and nighttime light intensity Robert C.M.Beyera, Sebastian Franco-Bedoyaa, Virgilio Galdo <https://www.sciencedirect.com/science/article/pii/S0305750X20304149#>
- (18) Beyer et al., 2018 Beyer, R.C., Chhabra, E., Galdo, V., & Rama, M. (2018). Measuring districts' monthly economic activity from outer space (No. 8523). World Bank Policy Research Working Paper. Google Scholar

[Home](#) > [Neural Computing and Applications](#) > Article

S.I: ML4BD_SHS | [Published: 27 May 2021](#)

An intelligent heart disease prediction system based on swarm-artificial neural network

[Sudarshan Nandy](#), [Mainak Adhikari](#), [Venki Balasubramanian](#), [Varun G. Menon](#) , [Xingwang Li](#) & [Muhammad Zakarya](#)

Neural Computing and Applications **35**, 14723–14737 (2023)

666 Accesses | **12** Citations | [Metrics](#)

Abstract

The accurate prediction of cardiovascular disease is an essential and challenging task to treat a patient efficiently before occurring a heart attack. In recent times, various intelligent healthcare frameworks have been designed with different machine learning and swarm optimization techniques for cardiovascular disease prediction. However, most of the existing strategies failed to achieve higher accuracy for cardiovascular disease prediction due to the lack of data-recognized techniques and proper prediction methodology. Motivated by the existing challenges, in this paper, we propose an intelligent healthcare framework for predicting

29. Jan MA, Khan F, Khan R, Mastorakis S, Menon VG, Watters P, Alazab M (2020) A lightweight mutual authentication and privacy-preservation scheme for intelligent wearable devices in industrial-CPS. *IEEE Trans Ind Inform* 1–11

30. Shynu PG, Menon VG, Kumar RL, Kadry S, Nam Y (2021) Blockchain-based secure healthcare application for diabetic-cardio disease prediction in fog computing. *IEEE Access* 9:45706–45720

Author information

Authors and Affiliations

Computer Science and Engineering, ASETK, Amity University Kolkata, Kolkata, India

Sudarshan Nandy

Mobile & Cloud Lab, Institute of Computer Science, University of Tartu, Tartu, Estonia

Mainak Adhikari

School of Science, Engineering and Information Technology, Federation University, Mount Helen, Australia

Venki Balasubramanian

Computer Science and Engineering, SCMS

School of Engineering and Technology,

Ernakulam, India

Varun G. Menon

**School of Physics and Electronic
Information Engineering, Henan
Polytechnic University, Jiaozuo, China**
Xingwang Li

**Department of Computer Science, Abdul
Wali Khan University, Mardan, Pakistan**
Muhammad Zakarya

Corresponding author

Correspondence to [Varun G. Menon](#).

Ethics declarations

Conflict of interest

The authors declare that there are no potential conflicts of interest in this work.

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

Nandy, S., Adhikari, M., Balasubramanian, V. *et al.* An intelligent heart disease prediction system based on swarm-artificial neural network. *Neural Comput & Applic*

35, 14723–14737 (2023). <https://doi.org/10.1007/s00521-021-06124-1>

Received	Accepted	Published
21 February 2021	11 May 2021	27 May 2021

Issue Date
July 2023

DOI
<https://doi.org/10.1007/s00521-021-06124-1>

Keywords

Artificial neural network

Heuristic formulation **Swarm optimization**

Back-propagation **Classification model**


Heart disease prediction

Exploiting Benefits of IRS in Wireless Powered NOMA Networks

Publisher: IEEE

Cite This

PDF

Xingwang Li  ; Zhen Xie  ; Zheng Chu  ; Varun G. Menon  ; Shahid Mumtaz  ; Jianh... [All Authors](#)

80
Cites in
Papers

1954
Full
Text Views



Abstract

Document
Sections

- I. Introduction
- II. System Model
- III. Sum
Throughput
Maximization
- IV. Numerical
Results
- V. Conclusion

[Authors](#)

[Figures](#)

[References](#)

[Citations](#)

[Keywords](#)

[Metrics](#)

Abstract:

We consider an intelligent reflecting surface (IRS) wireless-powered NOMA Internet-of-Things (IoT) network, where multiple NOMA IoT devices can communicate with access point (AP) with the aid of IRS. Specifically, IoT devices can harvest energy from a nearby appropriate power station (PS) during the wireless energy transfer (WET) phase through the direct links as well as the reflecting links from IRS. Then, the harvested energy can achieve the wireless information transfer (WIT) during the uplink between IoT devices and AP by adopting NOMA protocol. A novel resource allocation scheme is proposed to maximize the sum throughput of the considered system by jointly optimizing the time allocation factor and phase shift matrices of WET and WIT. The optimization problem is non-convex due to multiple variables. We first transfer this problem into convex function by assuming one of the phase shift matrices fixed, and solve the transferred problem by obtaining closed-form solution. Then, the elements collaborative approximate (ECA) and the manifold space gradient descent (MSGD) algorithms are designed to optimize the phase shift matrix. ECA is iteratively optimizing one element and MSGD is focus on the derivation of a gradient descent over the manifold space. Numerical results indicate the sum throughput enhancement by IRS compared with no IRS, and highlight the advantages of IRS.

Published in: IEEE Transactions on Green Communications and Networking (Volume: 6 , Issue: 1, March 2022)

Page(s): 175 - 186

DOI: 10.1109/TGCN.2022.3144744

Date of Publication: 20 January 2022  **Publisher:** IEEE

Electronic ISSN: 2473-2400

► **Funding Agency:**

I. Introduction

Internet-of-Things (IoT) technology is indispensable to the beyond fifth-generation (B5G) by delivering specified functions through connecting devices or "things" over the mobile Internet, which can greatly improve the access rate of mass devices [1]–[3]. The IoT has another advantage to connect a huge number of devices for information transfer and/or exchange [4]. As one of the core components of the IoT network, machine-type communication (MTC) devices are assumed to be maintained without

Expert Systems / Volume 39, Issue 5 / e12981

GUEST EDITORIAL | [Free Access](#)

Cognitive smart cities: Challenges and trending solutions

Varun G. Menon , Reza Khosravi, Alireza Jolfaei, Akshi Kumar, Vinod P

First published: 23 May 2022

<https://doi.org/10.1111/exsy.12981>

A smart city implies the realization of sustainable city growth enabled by technology-based intelligent solutions to give a good quality of life to its citizens. Information and communication technologies play a crucial role as the nerve centre of the smart city for collecting and analysing data from various sources, like mobile, social media, and sensors. Internet of things (IoT) and big data (BD) also play a critical role in the smart city infrastructures, changing the way we analyse patterns and trends in human behaviour. Smart cities generate a huge amount of data, and therefore need many flexible ways to implement data and processing gateways.

Recently, cognitive analytics have attracted the attention of researchers and practitioners worldwide as a technology-based smart solution. It is a novel approach to information discovery and decision making which uses multiple intelligent technologies such as statistical machine learning, deep learning, distributed artificial intelligence, natural language processing and visual pattern recognition to understand data and generate insights. A cognitive smart city refers to the convergence of emerging IoT and smart city technologies to realize cyber-physical social systems, their generated BD from sensing to communication and computing, and artificial intelligence techniques for all aspects of collaborative computing in sensors, actuators and human-machine interfaces. A cognitive city is one that learns and adapts its behaviour based on the past experiences and can sense, understand and respond to changes of a smart environment with many human and robotic elements. In cognitive cities, data flows not only from the citizens to city management systems (e.g., intelligent transportation systems and healthcare centres), but also from citizen to citizen. Citizens act as human sensors, and intelligence-enabled frameworks build a cyber-physical social system. Thus, consistent citizen engagement, ubiquitous data collection and sophisticated analytics are required to produce the best kind of cognitive city.

The implementation of cognitive smart city is highly context-dependent. Initiatives may range from incremental to disruptive and the deployment is shaped by many factors such as governance, economic, technology, social, environmental, legal and ethical issues. As smart cities projects become more pervasive across geography, technology and applications, it is imperative to identify key learnings to foster a deeper understanding of the technology evolution landscape and provide tangible benefits to smart city planners and key decision-makers. Viable intersection between technology solutions and digital urbanization design principles (people-centred and inclusive infrastructure, resilience and sustainability, interoperability and flexibility, managing risks and ensuring safety) need to be evaluated in order to provide balanced and replicable solutions.

In the research community, several works propose cognitive solutions that fit the needs of sustainable urban development. Academic literature, government consultation documents and policy papers articulate numerous challenges and research directions for incorporating cognition to realize new smart city services. Both qualitative and quantitative studies carefully consider BD analytics, semantic derivation and knowledge discovery, intelligent decision-making and on-demand service provision for a large number of smart city applications.

This special issue aims to stimulate discussion on the design, use and evaluation of self-correction and human cognition for continuous learning as key knowledge discovery drivers within socially connected urban ecosystems. The issue is focused on articles describing cognitive models for cyber–physical social systems with urban BD to leverage deeper insights from the vast amount of generated data with near real-time intelligence.

We received a very good response to our special issue call for articles. During the review process, each article was assigned to and reviewed by at least three experts in the field. After a rigorous multi-round review process, we were able to accept six excellent articles covering various topics related to cognitive smart cities. In the following, we will introduce those articles and highlight their main contributions.

The paper entitled, 'A taxonomy of energy optimization techniques for smart cities: architecture and future directions' discussed the need for devising more solutions for efficiently handling energy utilization associated challenges in smart cities. The paper presented a comprehensive survey on the energy optimization techniques in various systems, including the optimization techniques in block chain-based systems. Further, the paper presented a taxonomy that classifies energy optimization techniques and proposed an energy-efficient consensus mechanism, proof-of-high-performance optimization for high-performance computing-based ecosystems.

Authors in the paper, 'DFT: A deep feature-based semi-supervised collaborative training for vehicle recognition in smart cities' proposed a deep feature-based training (DFT) method for vehicle recognition in smart cities. DFT is also a semi-supervised collaborative training

method on the basis of two base learners. DFT adjusts data pre-processing and training process, optimizes the constructing a disagreement encoding network, and expands on the recognition disagreement of pseudo-labelled samples-based training sets. Compared with the typical collaborative training methods, DFT greatly accelerates the model's training process by reducing the convergence time, and improves the efficiency of vehicle recognition, while remaining the recognition accuracy unchanged.

In the paper, 'Sensor data fusion for the industrial artificial intelligence of things' the authors discussed a new framework for addressing the different challenges of the artificial intelligence of things (AIoT) applications. The proposed framework is an intelligent combination of multi-agent systems, knowledge graphs and deep learning. Deep learning architectures are used to create models from different sensor-based data. Multi-agent systems can be used for simulating the collective behaviours of the smart sensors using IoT settings. The communication among different agents is realized by integrating knowledge graphs. Different optimizers based on constraint satisfaction as well as evolutionary computation are also investigated in the paper.

The paper entitled 'A formal method for privacy-preservation in cognitive smart cities' presented a discussion on a technique for privacy-preservation in smart cities based on pseudonymization, clustering, anonymization and differential privacy methods. The modified clustering algorithm selects the initial cluster-based on the concept of dissimilarity between the data sequences. The paper also assessed the functional correctness and preformation of the proposed model for privacy-preservation in smart cities.

In the paper, 'Soft computing for abuse detection using cyber-physical and social BD in cognitive smart cities' authors discuss the use of soft computing techniques for abuse detection in the complex cyber-physical-social BD systems in cognitive smart cities. The objective of the paper is to define and identify the diverse concept of abuse and systematize techniques for automatic abuse detection for cyber abuse detection on social media and real-time abuse detection using social IoT. The cyber abuse studies on social media platforms have further been categorized as cyber-hate and cyberbullying whereas the real-time abuse includes studies of cyber-physical social systems. As in a cognitive smart city, citizens expect more from their urban environments with minimal intervention, this study helps us to establish the need to capture situational context and awareness in real-time and foster the need to develop a proactive as well as reactive safety mechanism to mitigate the risks of online abuse.

The paper entitled 'A lightweight intelligent intrusion detection system for industrial IoT using deep learning algorithms' discusses the techniques for improving intelligent decision-making actions in the industrial IoT (IIoT) network in a sustainable city. Main cybersecurity attacks are predicted by applying a deep learning model in the paper. The various security and integrity features such as Denial of Service (DoS), malevolent operation, data type probing, spying, scanning, intrusion detection, brute force, web attacks and wrong setup are

analysed and detected by a novel sparse evolutionary training mechanism-based prediction model.

We would like to express our sincere thanks to all the authors for submitting their papers and to the reviewers for their valuable comments and suggestions that significantly enhanced the quality of the articles. We are also grateful to Editor-in-Chief, Prof. Jon G. Hall and Special Issues & Reviews Editor Prof. Lucia Rapanotti for their great support throughout the whole review and publication process of this special issue, and, of course, all the editorial staff. We hope that this special issue will serve as a useful reference for researchers, scientists, engineers and academics in the field of cognitive smart cities' design and development.

[Download PDF](#)

About Wiley Online Library

[Privacy Policy](#)

[Terms of Use](#)

[About Cookies](#)

[Manage Cookies](#)

[Accessibility](#)

[Wiley Research DE&I Statement and Publishing Policies](#)

[Developing World Access](#)

[Help & Support](#)

[Contact Us](#)

[Training and Support](#)

[DMCA & Reporting Piracy](#)

[Opportunities](#)

[Subscription Agents](#)

[Advertisers & Corporate Partners](#)

[Connect with Wiley](#)

[The Wiley Network](#)

[Wiley Press Room](#)

Deep Reinforcement Learning for Intelligent Service Provisioning in Software-Defined Industrial Fog Networks

Publisher: IEEE

[Cite This](#)

[PDF](#)

Indranil Sarkar  ; Mainak Adhikari  ; Sanjay Kumar  ; Varun G. Menon  [All Authors](#)

5
Cites in
Papers

506
Full
Text Views



Abstract

Document Sections

- I. Introduction
- II. System Model and Problem Formulation
- III. Intelligent Service Provisioning Strategy
- IV. Performance Evaluation
- V. Conclusion

Abstract:

Fog computing has become a promising technology to improve the performance of low-powered Industrial Internet of Things (IIoT) devices by providing flexible and convenient computing services at the edge of the network with minimum delay. However, due to the ever-increasing traffic load in the network, the traditional service provisioning strategies can pose high complexity as well as network congestion, resulting in high energy consumption. Owing to this issue, in this article, we propose a deep reinforcement learning (DRL)-based service provisioning strategy in a software-defined industrial fog network to minimize the energy consumption of the network. The service provisioning strategy is performed in the network data plane, whereas the DRL is deployed in the control plane to enhance network efficiency. The service provisioning problem is formulated as the Markov decision process (MDP) and further solved by adopting the concept of a deep Q network (DQN). Further, we propose a task migration policy to ensure the high availability of computing devices while meeting a single point of failure (SPOF) issue. Finally, to show the effectiveness of the proposed method, it is compared with the traditional baseline algorithms over various performance metrics.

Authors

Figures

References

Citations

Keywords

Metrics

Published in: IEEE Internet of Things Journal (Volume: 9 , Issue: 18, 15 September 2022)

Page(s): 16953 - 16961

DOI: 10.1109/JIOT.2022.3142079

Date of Publication: 13 January 2022 ? **Publisher:** IEEE

▼ ISSN Information:

Electronic ISSN: 2327-4662

CD: 2372-2541

I. Introduction

A variety of delay-sensitive tasks have emerged with the exponential growth of the Internet of Things (IoT) applications, such as smart city, smart transportation, intelligent healthcare, etc. These tasks require significant computing resources for real-time task processing. Most of the IoT devices are embedded with sensors and are involved in applications that require either local data processing or offloading data to a suitable

Special Issue

Advances in Computational Intelligence Techniques for Next Generation Internet of Things




[View this Special Issue](#)

Research Article | Open Access

Volume 2022 | Article ID 7061617 | <https://doi.org/10.1155/2022/7061617>

[Show citation](#)

Towards Energy-Efficient and Delay-Optimized Opportunistic Routing in Underwater Acoustic Sensor Networks for IoUT Platforms: An Overview and New Suggestions

Varun G. Menon ^{1,2} Divya Midhunchakkaravarthy,³ Aaromal Sujith,² Sonali John,² Xingwang Li,⁴ and Mohammad R. Khosravi  ⁵

[Show more](#)

Academic Editor: Sumarga Kumar Sah Tyagi

Published: 17 Mar 2022

Abstract

In underwater acoustic sensor networks (UASNs), the reliable transfer of data from the source nodes located underwater to the destination nodes at the surface through the network of intermediate nodes is a significant challenge due to various unique characteristics of UASN such as continuous mobility of sensor nodes, increased propagation delay, restriction in energy, and heightened interference. Recently, the location-based opportunistic routing protocols seem to show potential by providing commendable quality of service (QoS) in the underwater environment. This study initially reviews all the latest location-based opportunistic routing protocols proposed for UASNs and discusses its possible limitations and challenges. Most of the existing works focus either on improving the QoS or on energy efficiency, and the few hybrid protocols that focus on both parameters are too complex with increased overhead and lack techniques to overcome communication voids. Further, this study proposes and discusses an easy-to-

with reduced delay. We simulate the protocol in Aqua-Sim, and the results obtained show better performance than existing protocols in terms of QoS and energy efficiency.

1. Introduction

The genesis of life on Earth had its inception on water, from which it went on to conquer varied frontiers. With the advent of the latest technologies, today's world is more connected than ever before, but ironically the blue planet still lacks efficient underwater connectivity. Underwater acoustic sensor networks (UASNs) [1–3] made their way into the limelight of research quite recently; its pivot objectives deal with an array of versatile interests from oceanographic studies dealing with marine geology, marine ecology, and physical and chemical oceanography. Another significant application is resource extraction, which mainly concerns harnessing abundant rare-earth minerals, petroleum, and natural gas under the sea bed, calamity prevention, deep-sea climate monitoring, and protection, surveillance, and reconnaissance of strategic waters by naval forces around the globe. Conventional methods used for undertaking these tasks mentioned above require humans to physically dive into the ocean's depths or rely on remotely operated underwater vehicles (ROUVs). After the data collection process gets over, these units resurface to provide the output. The data acquired always fell short of fulfilling its objectives as there were problems like the lack of accurate real-time data, stringent storage constraints, inability in handling mobility, and capability to withstand underwater pressure.

UASNs have cell-powered sensor nodes deployed throughout the ocean bed that interact with each other and with the sonobuoys located at the surface to suffice these objectives. Their presence ensures effective communication with the sensor nodes (real time), and they are also the first responders to notify the base station if any nodes fail. Besides all these conspicuous merits, they have been used in underwater acoustic research and antisubmarine warfare for a long time, reflecting its practicality as the UASN's function under tight frequency limitations. Numerous unique features of the underwater environment make the deployment and use of UASN quite a challenging task [4–7]. Underwater conditions are different from the situations on land where the communication takes place with radio frequency (RF) aid. Unfortunately, the underwater environment consumes the energy of the RF waves and renders itself impractical. The mobility of underwater sensors with the ocean currents is another major challenge. To get better off from the challenging underwater situations, UASNs communicate using acoustic waves [8–10]. Underwater acoustic waves typically operate in the frequency range of 10 Hz to 1 MHz. The delay of propagation accompanies this slender range, but it seems to be the only viable choice forward on modicum energy store. UASNs make another edge by facilitating interfaces to communicate with autonomous underwater vehicles (AUVs) remotely. This feature will exponentially increase the range of AUV control, and this merit will also help us perceive the underwater world for research with an added advantage of burgeoning the amount of ocean monitored by human beings, which currently accounts for only 5%. Routing of data packets right from the sensor nodes to the sonobuoys and to surface stations is one of the most challenging issues faced in UASNs, primarily due to the rapid energy drainage, limited bandwidth, significantly high latency, and reduced reliability [11–17]. Figure 1 presents a sample application scenario of UASNs.

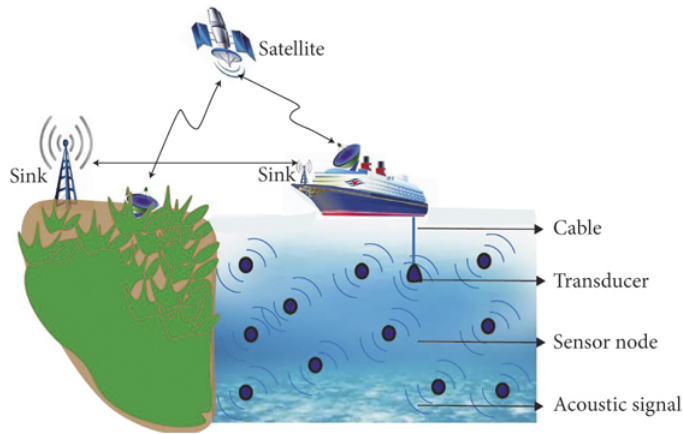


Figure 1

Underwater acoustic sensor networks (UASNs).

Terrestrial wireless sensor networks (TWSNs) have a conventional set of routing protocols that ensure good network performance. TWSNs at no point of operation face interruptions similar to ocean currents. Doppler spreading, interim path loss, and link quality loss create numerous challenges for routing in underwater environments. In a nutshell, the quality of service (QoS) and energy constraints of UASNs inextricably impede it from resorting to routing protocols of TWSNs. Majority of all the routing protocols proposed for TWSNs, thus proving to be powerless when it comes to UASNs. Numerous unconventional routing protocols were put forth in recent years for UASNs, and each focused on energy efficiency, thus improving various QoS parameters like throughput, latency, load balancing, and robustness. Some of these protocols have already been tested for research, military applications, and catastrophe prediction. The selection of an appropriate routing protocol is significant as it is answerable for the reliable deliverance of data packets to the destination.

Table 1 presents the variations between the terrestrial wireless sensor networks (TWSNs) and underwater acoustic sensor networks (UASNs). Routing protocols in UASN face numerous design challenges. The weightage given to path selection accounts for the various problems that have to be confronted in the underwater environment, such as marine aquatic life, acoustic disturbances, propagation delay, and seismic shadow zones. Many new routing protocols are proposed to tackle these dilemmas; however, most of them lack the description of appropriate routing strategies. Routing strategies advocate the parameters, which will be extensively useful for researchers and other professionals to calibrate the effectiveness of algorithms used in UASNs to develop a strategy to tackle limitations like high propagation delay and energy usage. Picking the suitable scheme ensures engineers achieve desired productivity in applications. The routing protocols for UASNs are mainly classified into location-based protocols and location-free protocols. The location-based protocols instrument the use of the information contained in the sensor nodes that are mostly two/three-dimensional position coordinates. In contrast, location-free/depth-based protocols depend mainly on pressure information present in sensor nodes. Most of the earlier conventional protocols proposed for UASNs selects the best path for sending data beforehand without considering the dynamic nature of the network environment. This negligence of the traditional routing protocols tends to compromise the use of the widespread resources in the network and can also pave the way to network failure. When setbacks like these started to portray, routing in UASNs

the demerit of unreliable transmission, that is, the undesired broadcast nature exhibited by the unreliable transmission is exploited here instead of selecting the nodes beforehand. The selection of nodes in ORP happens on the go. Numerous neighboring nodes (candidate set) receive the broadcasted message. The candidates belonging to the candidate set are sorted according to the metrics and prioritized based on the probability of becoming the next-hop forwarder. The candidate with the highest priority is given the ability to forward the data packets while others discard the packets. This is known as candidate coordination. Opportunistic routing protocols have proved their robustness and adaptability to uncertain conditions by showing their significant presence in many essential fields like oil/gas pipelines, power grids, and management of metro/railroads. Currently, the advancements closely related to ORP have not yet reached their pinnacle as many problems are yet to be solved. However, the prime intention that sleeps behind it is being the ability to make a set of independently weak nodes emerge together as a virtually robust set of links. Thus, ensuring reliability which in turn plummets the retransmission rates and chop down the energy consumption of UAWNs. All of these pros and cons will be thoroughly surveyed in this study.

Table 1

Differences between the TWSN and UASN.

The major contributions of the work are highlighted as follows:

- (i) We review all the major location-based opportunistic protocols proposed for routing data packets in underwater acoustic sensor networks over these years. Numerous energy-based, QoS-based, and hybrid location-based opportunistic routing protocols have been proposed in 2019, 2020, and 2021, and they promise to provide much better performance to various real-time applications deployed in UASNs. Very few works have provided reviews on these latest protocols, and we initially tried to address this research gap. We analyze and present a brief description of their working mechanism and highlight their issues and challenges. These issues can be taken up in the future for further improvement in the design of routing protocols in UASN.
- (ii) From the review of the latest protocols, it is observed that increased delay and energy drainage are the two significant areas of concern that need further solutions. We then tried to model an easy-to-implement routing protocol that can guarantee energy efficiency with reduced delay to various applications deployed in UASNs.

The rest of the study is organized as follows: in Section 2, various applications using underwater acoustic sensor networks are discussed. The fundamental principles of opportunistic routing are explained in Section 3. Energy-based, QoS-based, and hybrid location-based protocols are discussed in Section 4. The proposed energy-efficient and delay optimized protocol is discussed in detail in Section 5. The future research directions are discussed in Section 6, and the study concludes in the next section.

2. Underwater Acoustic Sensor Networks

The planet we dwell upon is covered approximately with 71% of water. Under this blue element, lies countless untapped resources that will enable human society to advance in countless ways. In order to consolidate this final frontier, underwater wireless sensor networks prove to be the need of the hour. The underwater wireless sensor network (UASN) is the collection of self-driven sensor nodes and

data to the target destination using acoustic signals. The attractive applications of UASNs comprise real-time surveillance, disaster prevention, navigation assistance, water quality determination, industrial organization, fish tillage, underwater exploration, and pollution tracking. Almost each of these applications of UASN demands sensor nodes to transfer sensed data timely and precisely through the source node present underwater towards the destination node on the surface with the help of intermediate nodes in the network. However, due to the dynamic nature of UASNs, continuous node mobility, communication voids, and limited battery storage often lead to degraded network performance. In this complicated underwater environment, how to route data packets promptly and effectively even with the presence of a communication void is the most challenging research question.

2.1. Challenges in UASNs

This section presents an overview of various challenges in underwater acoustic sensor networks.

2.1.1. Acoustic Communication

The terrestrial networks depend on RF waves to execute communication. Whereas the underwater environment is highly unpredictable, and RF waves are unfortunately absorbed. Additionally, a high amount of attenuation paves the way to energy loss; hence, RF waves are not an option to be considered. Optical waves cannot be regarded as a choice because the mobility of the nodes is unable to guarantee accuracy. The only viable option in this scenario is acoustic waves. Many other flaws are associated with the usage, but a suitable routing protocol is expected to sort out the dilemma.

2.1.2. High Mobility

The underwater sensor nodes are constantly on the move. These movements are caused by ocean currents which arise due to wind, breaking waves, temperature, and salinity variations. For efficient data gathering, the movement of these nodes is vital and indispensable. However, in reality, this high mobility induces the formation of curves to the acoustic waves, which triggers the emergence of zones that makes some of the sensor nodes in the network unable to participate in data transfer, which brings forth performance issues to the network.

2.1.3. Underwater Noise and Interference

The ocean is packed with a lot of noises and interferences that arise from varied sources. Some of the underwater noises are caused by breaking waves, rain, and marine life. However, various sources are man-made like, shipping, military sonars, fishing, and research activities. These disturbances affect the quality of data packet delivery in the underwater environment.

2.1.4. Low Bandwidth

The operational frequency range of the underwater sensor nodes is primarily restricted due to the usage of acoustic waves as the medium of transmission. The bandwidth is a meager spectrum that lies between 1 kHz and 50 kHz. This poses a significant problem for routing protocols as it requires an enormous amount of data exchange at various stages like discovery and maintenance. Tight bandwidth constraints

dynamic nature of UASNs, continuous node mobility, communication voids, and limited battery storage often lead to degraded network performance. In this complicated underwater environment, how to route data packets promptly and effectively even with the presence of a communication void is the most challenging research question.

2.1.5. Low Data Rate

Speed is a crucial factor when it comes to information exchange. The faster the data reaches the destination, the better. Unlike its counterpart (terrestrial environment), the speed at which data is transmitted in the underwater environment is influenced by numerous factors. Firstly, the propagation speed of acoustic waves is inferior to RF waves by many folds that create room for propagation delay. Secondly, there are various persuasive components like depth, temperature, and the degree of salinity of the water. The data rate is meager and accounts for approximately 100 kbps or occasionally a bit more.

2.1.6. Transmission Loss

The hurdles when it comes to underwater sensor network implementation are numerous. Acoustic waves do not guarantee any reliability for the network. On top of that, path loss, Doppler spreading, and high latency will provide a clear picture that there is a considerable amount of packet loss. Transmission loss in any network is not a desirable outcome. Interference is one of the main reasons for packet loss to occur. When the same nodes participate in data transmission continuously for an extended period, the battery can run out, resulting in a communication blackout, and the node will no longer be able to transfer data collected from some places in the network. The acoustic signals have open channels, which are more likely to be utilized by an attacker or malware and to wreak havoc in critical services like routing, localization, and synchronization. Delay variance and bit error are two constituents that can lead to a high amount of packet loss and bit error rates.

2.1.7. Error Prone

The underwater sensor nodes, unlike their conventional counterparts, are not reliable. The environment in which it has been implemented does not support its operation. The mobility, high latency, delay in propagation, high interference, noise, etc., make them highly susceptible to errors. The changes that manifest due to the variation in salinity, depth, and acoustic speed have an implicit effect on making the data transmission process error prone.

2.1.8. High Energy Consumption

Energy drainage is a significant problem in UASNs. Communication between various nodes in a network rudimentarily requires the acknowledgment of its position. The nodes are constantly swaying in harmony with the ocean currents, and it is essential to update their position consistently with their neighbors for effective participation in the data transmission process. Ironically, this position-update process drains quite some energy from these sensor nodes. Another avenue wherein the power consumption rates rocket is when packets have to be retransmitted due to high interference. The data load bestowed on end nodes that connects to the surface stations rapidly depletes the battery power, resulting in the termination of connection with the rest of the nodes in the underwater network. The consumption rates vary with the depth in which it is implemented: the battery is expected to operate well in shallow and deep-water

2.1.9. Channel Attenuation

Channel attenuation is another dilemma that underwater sensor nodes have to confront. The implementation of sensor nodes in the ocean bed is beneficial only if it can collect and transfer data, but due to channel attenuation, the data collected cannot be efficiently extracted from the received signal.

2.1.10. Short Network Lifetime

The pivot grounds for the short lifetime of underwater sensor nodes are due to its source of energy, the battery storage. The nodes extensively consume energy while localization, routing, and data transfer. The hostile underwater condition makes the replacement of cells regularly a challenging task. Therefore, an efficient routing protocol has to consider energy consumption while making decisions on routing path selection. Furthermore, if the nodes run out of energy, it will result in the formation of dead nodes that can affect the network performance and data transmission to the surface stations.

2.1.11. Security and Privacy Issues

UAWNs are made and deployed to monitor places consistently that are far from the shoreline. There is a high chance that the nodes are deployed in strategic waters for specific applications. Attackers can easily manipulate UASNs to inject malicious attacks upon the network. They can also be physically destroyed by enemy divers/AUVs. In the worst-case scenario, attackers can inject fake nodes into the network to provide misguided information and use compromised nodes to extract exclusive data from the network.

2.1.12. Physical Challenges

The sensor nodes have to be fabricated so that they are compact, rigid, and waterproof and should also be able to withstand the pressure of water. Marine life is the next physical challenge that these sensor nodes have to face. In reality, it is physically impossible to protect every one of the nodes in the network.

2.2. Applications of UASNs

This section presents an overview of the major applications that use UASNs.

2.2.1. Military Applications

The military applications of UWSNs can cover a wide range of requirements from monitoring to reconnaissance. In 1982, the United Nations Convention On the Law of the Sea allowed countries to exercise jurisdiction on territorial waters up to 200 nautical miles along the baseline. The naval force guards the coastline against invaders, but the underwater regions in strategically important areas are vulnerable. This vulnerability can be defeated with the help of UASNs. It will enable the militaries to detect enemy divers, submarines, torpedoes, AUVs, and naval mines. The ability to get real-time data will enhance strategic decision-making.

2.2.2. Oceanography

in the ocean and also be able to predict or artificially simulate similar conditions that will benefit human society. Oceanography can provide efficient analysis if it can collect real-time data. UASNs can be utilized to perform experiments to unravel the mysteries of the underwater world consistently, which will indirectly help us to obtain solutions to various problems oceanographic problems.

2.2.3. Coral Conservation

Corals are one of the most beautiful living things on the planet. It takes millions of years to form barrier reefs. The Great Barrier Reef in Australia is the most extensive collection of corals on the phase of the Earth. However, it is dying due to coral bleaching due to the shift in climatic conditions throughout the world. Coral reefs around the globe are on the verge of extinction, and nations are trying to change the situation. Human interference has the likeliness to accelerate the degradation process, but data collection is seemingly impossible without human interference. UASNs are a profound solution that can provide real-time data to conservationists by limiting human interference.

2.2.4. Resource Tapping

The Earth has resources that are tucked away in the depth of the ocean. UASNs provide a way to get the know-how of these varied resources. UASNs will enable us to find out the location, approximate quantity, and dispersion pattern of resources present in the ocean bed. Petroleum and natural gases are an inevitable part of our civilization, which are unfortunately limited. New potential sources can be located for extraction using UASNs.

2.2.5. Fishing, Farming, and Recreation

The fishing industry will benefit from the use of UASNs as they will help locate groups of fish. Underwater farming has been used to cultivate seaweed, lettuce, basil, etc., in countries like Japan and Italy. Knowing the nature of the ocean is an integral part of taking out cruises for recreation purposes deep into the ocean. UASNs can tell leisure seekers about the risks of a tsunami or hurricane before they embark on a cruise.

2.2.6. Disaster Prevention/Prediction

The implementation of UASNs will enable us to detect in advance any underwater earthquakes and volcanic eruptions, which will help us prevent or predict disasters. Many aircraft have gone missing over these years in oceans, and no data about their disappearance was harnessed. One of such shocking incidents is that of Malaysia Airlines 370. UASNs, provided it is implemented correctly, will enhance us to chart the ocean, and it is possible to derive a pinpoint location of any possible crash site.

2.2.7. Climate Change

The rising sea levels and warmth have grabbed international attention. The polar ice caps are at the risk of meltdown. Accurate screening and reports of the polar meltdown can be undertaken with the help of UASNs, which will help researchers and environmentalists to find solutions to these baffling dilemmas.

The uncertainty of the underwater environment poses many underlying threats to establishing efficient communication strategies. The constraints on power and the constant movement of nodes due to tides make opportunistic routing (OR) a viable solution. The opportunistic routing owns different modus operandi. On receiving a data packet, the host node takes into consideration a set of eligible neighboring nodes and prioritizes them based on various parameters. These parameters are different facets like the node closest to the destination and least power draining. A suitable packet forwarder gets opted from the candidate set based on priority and availability. Thus, in the case of unreliable underwater communication, OR proves to be promising as it provides extended reliability, robustness, and QoS than other legacy routing methods.

The principle of opportunistic routing idea was initially developed in ExOR [19] in 2005. The notable advantage of this protocol is that this protocol exploits the multiplex communication opportunities in which the broadcast character belonging to the wireless network develops. The fundamental working of opportunistic routing can be implemented in this protocol, and the three main steps include the following: Initially, the sender node can broadcast the message data. Secondly, upon receiving that data, one relay node is selected as the best forwarder node. After that, the selected best relay node transports the message transmission to the next best relay, and so on. The method is continued until the data reaches the target position. Compared to traditional routing methods, the next-hop relay is selected only after it has received the data, thereby reducing the number of data retransmissions.

The UASN's operation using acoustic channels for communication has many downfalls as there is a prevalence of solid attenuation, time-varying multipath, ambient noise, and modicum propagation speed. All of these contribute to increased delay, error, limited bandwidth, high energy consumption, communication cost, and at times temporary loss of connectivity among nodes of the network. The profound influence of channel fading is crucial to confront as it can directly impact declining routing performance. The scope of application of OR is of paramount importance because it has significantly low retransmission rates, in turn reducing the power consumption. Assured packet delivery facilitation by opportunistic routing ensures no wastage of network resources. It also reduces the chance of system collapse and diminishes retransmission costs. Additionally, opportunistic routing is a versatile choice as it applies to a variety of networks.

Opportunistic routing facilitates a dynamic and instant multiple-path routing technique through opportunistic relay selection, unlike the traditional routing method. Instead of a single precomputed relay, opportunistic routing initially broadcasts a data message to a set of forwarder relay nodes. Fundamentally, these forwarders are organized according to a particular unit. The idea of this method is to select the best forwarder relay among all the nodes. This selected relay manages the process of forwarding the packets. These steps are recursively executed until the packet is transmitted to the target destination. The rudimentary opportunistic routing mechanism consists of the following four steps:

- (i) Forwarder relay set choosing
- (ii) Broadcasting of data to forwarder nodes
- (iii) Coordination scheme is used for best relay selection
- (iv) Forwarding of data by the best relay

Each node in opportunistic routing broadcasts a data packet to various adjacent hops periodically. Hence, if communication to one neighbor crashes, another nearby node that has received the data packet can transmit it. OR defines a group of various next hops as the best forwarder relay collection and indicates it as a forwarder relay set (FRS). When a message gets transmitted to FRS, numerous forwarders can obtain a similar message packet. We can avoid duplicate transmission by selecting a single candidate as the best

transmits that to the target. Otherwise, the node that has the next most significant priority transmits the data packet, and so forth. The leftover candidates will discard this data packet. The FRS selection is divided into three main components as follows: (a) forwarder relay discovery, (b) prioritization variable calculation, and (c) forwarder relays selection, prioritization, and filtering.

To find out the neighbor node, periodic or nonperiodic packets are broadcasted. This neighbor node depends on the link quality, which is changeable as well as dynamic. Hence, this phase is given the charge of computing the stability and quality of the links to reach the neighborhood. Based on these values, a group of nodes is determined. Initially, every node located in the vicinity of the sender node is added to the FRS. Then, the forwarders are taken and sorted based on the chosen variable. After an FRS is elected, priorities are given to forwarder relays based on a specific value. So, the variable selection affects the network throughput significantly. The priority variable election depends on the routing application needs and targets. For some applications, such as emergency recovery, the position data is essential. Consequently, the relay nodes must know their location, and routing can be executed by selecting the relay that is locationally nearest to the target node. Controlling the number of forwarder relays can reduce the overhead and duplicate data transmissions. Moreover, since the size of FRS grows, the number of forwarders who cannot listen to one another also increases. This leads to duplicate packet transmissions. Hence, it is suitable to avoid some forwarder candidates from the FRS. This approach is termed candidate filtering or forwarder. The conventionally used filtering method avoids the forwarders, which are not suited instead of the source. But somehow, this policy cannot guarantee efficient network performance. Another method of FRS creation is discussed in some previous works. The technique is based on implementing an algorithm that can optimize correctly and compute the optimal forwarder relays set for every node, such as forwarder relay sets that are created by Dijkstra's algorithm. However, these methods do not solve the issues of duplicate data transmission.

The optimal relay selection that uses a coordination scheme is used for the coordination of packet forwarding operation between next near nodes. This scheme is responsible for selecting the most suitable forwarder relay to push forward the data packet. Coordination methods need signaling between forwarders. The basic coordination schemes are generally classified into a timer-based, contention-based method, and token-based coordination scheme. In a contention-based method, the main principle is that forwarder relays contend to transmit the data packet with the help of control messages. For example, if a sender node transmits a forward request, its near hop nodes have to compete with themselves to come to a consensus on the forwarding of the data packets. In the timer coordination method, the forwarder relays are supposed to be ranked according to a specific priority value. This rank is commonly added with the message header, which is consistent with the hierarchy in which potential forwarders are permitted to respond. Thus, the largest priority node is allowed to respond to the first slot. The next priority node responds to the upcoming time slot, etc. However, this method is straightforward and easy to carry out, timer-based coordination that incurs some delay, affecting network performance. Another method is the token-based coordination scheme, in which the transferring of data packets is only possible through a token holder. In this scheme, the duplicate message transmission is completely prevented, but it faces increased overhead control. A forwarder (relay) node contains the overhead data packets that are being sent when a token arrives. Tokens travel with connected forwarders because more miniature priority forwarder relays can listen to large priority nodes. If no token arrives, the candidates may be moved into an idle state, slowing down the network.

4. Location-Based Opportunistic Routing Protocols in UWSNs

All routing protocols for UAWNs can be classified as location-based protocols and location-free protocols. The location-based protocols instruments data contained within sensor nodes that are mainly

forwarding (VBF) [21] was one of the earliest protocols proposed in this category. This protocol can fabricate some virtual vector pipe that exists between the source and the destination. Only the nodes in the vicinity of the “vector” right through the source and destination will have the ability to do message forwarding. Hence, routing involves only a tiny group of nodes.

Similarly, numerous location-based protocols were proposed for UASNs. Some of the protocols focused on improving energy efficiency, while others focused on improving QoS parameters like delay and delivery ratio. Recently, many hybrid protocols also have been proposed which consider both energy efficiency and QoS. This section presents a comprehensive discussion on all the latest location-based protocols proposed for UASNs.

VBF makes use of the node location information to make the routing decisions. The knowledge of the position of nodes encourages it to be faster, reliable, and scalable. With the protocol, a virtualized pipe is created from source to destination, and those nodes in the pipe possess a higher probability of becoming the forwarder nodes, while the nodes outside are disregarded. Sink-initiated query and source-initiated query are the two main ways in which VBF addresses routes to different queries. Conceptually, all nodes inside the virtual pipe have the eligibility to forward the packets, but due to limitations like energy, mobility, and propagation delay of acoustic waves, a self-adaptation algorithm was suggested. Another protocol, directional flooding-based routing (DFR) [22] defines a forwarding method formulated by the angle among the center and intermediate nodes. The nodes are responsible for forwarding the data packets through a flooding method. It also considers the quality of the link between the sender and the destination node. A significant concern with this protocol is redundant data transmission and increased energy consumption. The information-carrying routing protocol (ICRP) [23] is an influential conservative, continuous, and versatile directing protocol. The sender hub checks the current location to the final destination when it owns the data to be sent. If there is no current course or path, it starts a path development process by communicating with the information packet, conveying the route disclosure message. Every node present on the network communicates to maintain the reverse route with the information path.

Hop-by-hop vector-based forwarding (HH-VBF) [24] is another variation of VBF where each forwarder resorts to a different routing vector. HH-VBF rudimentarily is just a version of the vector-based forwarding protocol. HH-VBF is a viable option compared to VBF as it can work well with sparse networks and is not liable to the routing pipe radius threshold. However, there is an increase in the computational delay, which in turn degrades the network performance. Reliable and energy balanced routing (REBAR) protocol [25] is a routing protocol that is energy efficient that helps in varying the broadcast domain. REBAR has good reliability and increased lifespan of the network. To balance the energy consumption in the network, a flexible scheme is developed to establish the data propagation range. Here, the nodes near the destination have a modicum radius. However, the increased node movements may expend an excessive amount of energy, resulting in degradation of performance. Vector-based void avoidance (VBVA) [26] protocol is simply an extension of the VBF protocol and focuses on addressing the void problem and energy efficiency in UASNs. The protocol works similarly to VBF when there is no void but uses a revised strategy when voids appear in the network. This helps the protocol to maintain better energy efficiency even with voids in the network.

The energy-efficient and collision aware (EECA) [27] multiple-path routing methods are founded on computing two different collision-free paths using restricted-energy modified flooding. It is one of the earliest protocols that gave equal importance to the betterment of QoS and energy efficiency in the network. Here, multipath power-control transmission (MPT) allows packet data transfer within limited end-to-end data error value and reduced power of transmission. The reliable energy-efficient routing protocol [28] functions on the foundation of link quality, physical data distance, and energy available in the UASN. These three metrics are calculated and shared with all nodes in the network. The protocol uses

Location-aware routing protocol (LARP) [29], the GPS is used to identify the exact area of the sink nodes. The sink nodes then broadcast the location information in the network. At the least three sink nodes are used for reference, other nodes in the network calculate their position. The sender can locate the next hop by broadcasting two things as follows: (1) location of the destination node and (2) moving direction of the packet. Packets are forwarded if the receiving node discovers that it is moving in a similar direction. The quality-of-service aware directional flooding-based routing (QoSDFR) [30] extends the DFR protocol. In this routing strategy, the sink node is responsible for sending feedback to various other nodes in the network about the channel condition, and based on the feedback, the optimal forwarder is selected. Protocol results in high throughput because of the limited energy consumption and varying channel conditions.

Scalable and efficient data gathering (SEDG) [31] protocol tries to increase the delivery ratio of the packet and also saves the modicum energy by feasible assignment of the member nodes and gateway node (GN). Here, an autonomous underwater vehicle (AUV) goes through the network area with a precomputed elliptical route and collects data from the gateway node (GN). AUV-aided efficient data gathering (AEDG) routing protocol [32] employs an AUV to gather information from gateways or intermediate nodes and use the shortest path tree (SPT) algorithm to balance the energy consumption. Besides that, AEGD designs a model that improves the result and saves energy by reducing the node members. Furthermore, the nodes live for an extended time to transfer data, thereby increasing delivery chances. The delay-aware energy-efficient routing protocol (DEEP) [33] is a delay-aware routing protocol dependent on collision rate and energy. DEEP makes use of an adaptable node aimed to minimize the collision rate. All the intermediate nodes are elected by virtue of delivery ratio and link quality. In the channel aware routing protocol (CARP) [34], the next-hop transmitter node is elected due to its distance from the previous intermediate node and available energy. Here, every intermediate node is familiar with its neighborhood between the destination node and the next hop. Sender then broadcasts a PING message to the network to compute the next forwarder. Considering a case where the hop value of a sink is lower than the sender node, it replies a PONG data. CARP uses an efficient relay selection method, which doubles the packet delivery ratio.

The novel efficiency forwarding protocol (NEFP) [35] is a proactive anycast routing protocol proposed for UWSNs. It promotes three different approaches. One defines a routing method that avoids unnecessary forwarding of packets where the collision dilemma is averted using a timer. Moreover, finally, the design uses Markov chains to calculate the probability of forwarding the data packets that encourages adaptability to constantly changing network topology. Nevertheless, the performance of the suggested protocol is decreased in the sparse region and as a result, reduces the number of forwards in the phases. Geographic and opportunistic routing protocol with depth adjustment (GEDAR) [36] is a geo-opportunistic routing protocol proposed for a minute-monitoring task. It utilizes a greedy forwarding method to advance the message towards the next hop. The source node chooses the best candidate from the forwarding set. The opportunistic routing in GEDAR reduces the number of retransmissions. GEDAR uses a recovery mode that helps to avoid the void areas. If a node is present in the void area, it will adjust its depth to overcome the void, and new messages will be queued. The greedy strategy will reschedule the node later. Markov model-based routing (MMVR) [37] selects its route from the lower surface to the top level based on changing data traffic. The routes are stable and adaptable, with fewer hops from the sender node and destination. In a localization-based dynamic routing protocol (LBDR) [38], the network is split into smaller layers, and a virtual routing vector is made within the sub-layers. Nodes will move in and out of the virtual vector based on the water current, resulting in high throughput. Void handling geo-opportunistic routing (VHGOR) [39] protocol focuses more on efficiently handling the communication holes in the network. Here, a quick hull algorithm is used to avoid a convex hull. When the node or hub reaches a convex region, rebuilding the convex void helps check an alternate and different way to resume the greedy transmission. VHGOR improves the network performance in networks with voids compared to

Geographic and opportunistic routing (GOR) [40] protocol shows efficient multi-hop data transmission in UWSNs with an upgraded strategy compared to previous protocols. Sometimes, this method gives room for the formation of the void region, and GOR tackles this issue using some void-handling algorithms. The framework considers network density, traffic load, and energy control features to bypass the empty region. The range-based low overhead localization technique (LOTUS) [41] significantly improves on earlier versions of the localization protocols. The protocol can estimate locations based on only two references, enabling this technique to work in networks with fewer nodes. The geographical duplicate reduction flooding (GDflood) [42] considers the location data regarding sensor nodes and joins it with network coding. Energy-efficient grid routing based on 3D cubes (EGRCs) [43] employs a 3D cube network that is subdivided into small cubic clusters. The cluster head is determined based on the remaining energy and position of the intermediate node. All the cluster heads then compute their intermediate node based upon the delay and location. EGRCs reduce energy consumption and end-to-end delay and increase the network performance.

Mobile energy-efficient square routing (MEES) [44] is a routing protocol focusing on energy efficiency in underwater sensor networks. The method uses a division of the network field into dense and sparse regions. A major advantage of this method is that, the mobile sink shifts in a clockwise direction that ensures the highest coverage of nodes in the network which will, in turn, result in high throughput and energy consumption. Topology control vector-based forwarding (TC-VBF) [45] is a revamped version of VBF, which tries to address the limitation of VBF in light conditions. Another protocol energy-efficient multipath grid-based geographic routing protocol (EMGGR) [46] fragments the network into 3D grids. The routing is executed in a grid-by-grid fashion with the help of gateway nodes. Disjoint paths result in high energy efficiency and a good packet delivery ratio. Balanced multiobjective optimized opportunistic routing (BMOOR) protocol [47] uses a strategy where the data from the lower surface takes the best route through the intermediate nodes to the top-level sink. Here, the nodes are located as per dynamic assessment with regards to optimal energy forwarders. The BMOOR protocol needs no spatial data, which is costly in UWSN. The protocol is developed using a generation-based bio-inspired, meta-heuristic algorithm. This helps in delay depreciation and maximization of delivery ratio, and thereby the network lifetime is enhanced. Another proposed protocol for UWSN is energy-efficient interference aware routing (EEIAR) [48] that opts for the best forwarder following the shortest distance. The shortest distance determination decreases the propagation delay. The power control-based sharp directing routing (PCR) [49] selects the most optimal transmission power level available at each submerged sensor node, which helps improve the packet delivery conveyance at each round. Also, it condemns the usage of high-power transmission and the uncontrolled consideration of neighboring hubs in the following hop candidate set, which would end up being the root cause for building the energy utilization on the network. The simulation outcomes depict that PCR diminishes the energy expenditure by adjusting the transmission power and electing the best candidates. The stateless opportunistic routing protocol (SORP) [50] uses a novel method to employ a variable forwarding area that can be reshaped and replaced according to the regional density and placement of the potential forwarding nodes to improve the energy and reliability. The protocol gives good performance compared to the previous protocols. Glider-assisted link disruption restoration mechanism (GALDRM) [51] uses a link disordering recognition with a related link rebuilding method. In the connection acknowledgment system, the group nodes gather the link data. The cluster heads gather the disruption data in link disruption and then schedules gliders as relay nodes to revive the link. Utility capacity is built up by limiting the channel. A multiplier technique illuminates the ideal area of a lightweight flyer. The simulation outputs exhibit a glider-assisted reconditioning procedure that helps to reduce energy consumption. The energy-aware void-avoidable routing protocol (EAVARP) [52] expands the network lifetime and packet delivery rate in underwater sensor networks. EAVARP includes layering and data collection phase with the help of directional forwarding strategy and uses residual energy and data transmission to avoid cyclic transmission and flooding. Fuzzy logic-based VBF

vicinity of the target node. The projection angle allows it to be selected onto the virtual routing vector pipe. The best advantage of this protocol is that it achieves better energy and throughput. However, nodes in the selected vector terminate on dealing with a high load of the message, which is similar to the conditions in VBF.

Mobility-assisted geo-opportunistic routing (MSAGOR) [54] protocol is mainly based on interference avoidance. Here, the network region is fragmented into compact cubes to diminish the interference, which helps to make additional well-informed routing strategies for better energy utilization. Moreover, an optimal number of transmitting nodes are selected from each cube based on its distance to the destination. This proximity will help to avoid void nodes. The extensive simulation results reveal that this protocol will maximize the delivery ratio and network lifetime. Totally, opportunistic routing algorithm (TORA) [55] is an anycast, geographic opportunistic routing protocol proposed for UWSN. The protocol is implemented to avert parallel transmission, bring down end-to-end delay in the network, tame the dilemma of void regions, and enhance network throughput. TORA uses time on arrival and its range-based equation to localize nodes. The energy-aware opportunistic routing (EnOR) [56] is an energy-aware opportunistic routing (EnOR) protocol that can adjust the priority level of forwarding between candidate nodes. This leads to steady energy utilization and increased network lifetime. By using the residual energy, link reliability, packet advancement ratio, and EnOR change the priority of transmission level. Adaptive hop-by-hop cone vector-based forwarding protocol [57] tries to improve the reliability of data transmissions in the sparse sensor regions by making some modifications to the base angle of the cone as per the network structure. These protocols improve the network performance by reducing the number of duplicate packets and also enable a better selection of the potential forwarder node.

Authors in reference [58] discuss implementing a modified strategy for depth-based routing that can transfer the data reliably to the surface sonobuoys. The technique mainly uses the 2-hop neighbor technique and tries to improve the delivery ratio of packets in the network. Authors in reference [59] proposed a technique combining the ant colony optimization algorithm, artificial fish swarm algorithm, and dynamic coded cooperation to improve efficiency by reducing energy consumption. Improving the flexibility of the protocol with the network was one of the major tasks of the proposed algorithm, along with finding the most optimal route. In reference [60], authors presented a Q-learning-based multi-hop cooperative routing protocol for underwater networks. Using this algorithm, the nodes with maximum Q-value were selected as the next forwarders in the network to transfer data from the source to the destination. A coding-aware strategy was proposed for efficient routing in networks with the sparse deployment of nodes [61]. The topological information was used to expand the candidate set using the protocol. An interesting approach that utilizes AUVs to carry sensor nodes to repair the routing voids when foreseeing the occurrence of voids was proposed in reference [62]. The protocol initially predicted the location for repair and then directed the AUVs to the particular location to carry out the repair process. Most of the proposed protocols are complex and incur high overhead, which degrades the performance of the network. Although many of the current protocols improve the data delivery ratio significantly, it comes at the cost of increased energy consumption. It is vital to develop a simple to implement a protocol that can take care of energy efficiency in the network while ensuring reduced delay in the network.

5. Energy-Efficient Location-Based Opportunistic Routing Protocol (EELORP)

In this section, we present the discussion on the proposed energy-efficient location-based opportunistic routing protocol (EELORP) that is designed to provide better energy efficiency and data delivery with

Initially, we try to provide a theoretical analysis to the proposed protocol. The focus is mainly on the delay of transmissions that can be reduced further to enhance the performance of the system. In the underwater network, a delay occurs within two different links, the wireless sensor to the wireless controller and wireless controller to the actuators. The delays are denoted by $T_{w(s-c)}$ and $T_{w(c-a)}$. Assuming the controller to be time invariant, the delays due to two sources are combined together to get total wireless sensor network delay as follows:

$$T_{wt} = T_w(s - c) + T_w(c - a). \quad (1)$$

The computation delay of the controller can also be included in the total wireless sensor network delay. As the assumption in the wireless controller is time invariant, the decision of controller $d(t)$ is independent of the time it receives the sample $S(\gamma b)$. So, the total wireless sensor network delay is only important for us. The analysis of UWSN stability is carried out by assuming two different scenarios as follows: (a) the continuous UWSN network delay system is considered by determining UWSN stability with constant network delay and (b) the discrete UWSN network delay system is considered by determining UWSN stability with time-varying networking delay.

For a continuous UWSN network-delayed system, the UWSN having total network delay as T_{WT} at the time $t = \gamma b$ is considered. The assumption is extended by making $T_{WT} < b$ for all values of “ γ ” belonging to “ S ” with $\gamma \in s$. The system is modeled mathematically as follows:

$$\frac{\partial}{\partial t} l(t) = Pl(t) + Qd(t - T_{WT}), \frac{\partial}{\partial t} l(t) = Pl(t) + Qd(t - T_{WT}), \quad (2)$$

where “ t ” belongs to $[\gamma b, \gamma b + b]$. Also with P , Q , R , and S as known matrices, we have the following relation:

$$M(t) = R\gamma(t) + Sd(t). \quad (3)$$

$d(t)$ is the received signal with no delay and $d(t - T_{WT})$ is the received signal with delay. In the case of $d(t)$, that is, received signal with no delay $d(t) = d(\gamma b)$ for $t \in [\gamma b, (\gamma b + b)]$.

Proposition 1. *The UWSN with the delay mentioned above validates the below-mentioned difference equations. The derived equation is as follows:*

$$l(\gamma b + b) = \tau l(\gamma b) + \varepsilon_0(T_{WT})d(\gamma b) + \varepsilon_1(T_{WT})(d(\gamma b - b)). \quad (4)$$

For $\tau = \int_0^b e^{Pb}$, we obtain the following equation:

$$\varepsilon_0(T_{WT}) = \int_0^{b-T_{WT}} e^{Pb} d\gamma\varphi, \quad (5)$$

$$\varepsilon_1(T_{WT}) = \int_{b-T_{WT}}^b e^{Pb} d\gamma\varphi.$$

Now, we have the following equation:

$$l(t) = e^{P(t)} + \int_0^t P(t-r)\varphi d(\gamma) d\gamma. l(t) = e^{P(t)} + \int_0^t P(t-r)\varphi d(\gamma) d\gamma. \quad (6)$$

If the delay $t_0 > 0$, then the above equation is rewritten as follows:

$$l(t) = e^{P(t-t_0)} l(t_0) \quad (7)$$

For for $t > t_0$, we have the following equation:

$$l(\gamma b + b) = \tau l(\gamma b) + \epsilon d(\gamma b), \quad (8)$$

where $\tau = e^{Pb}$ and $\epsilon = \int_0^b e^{Pr} d\gamma \varphi$.

Applying equation (7) to (4), we obtain the following equation:

$$l(\gamma) = e^{Pb} l(\gamma b) + \int_{\gamma b}^{\gamma b + b} e^{P(\gamma b + b - \gamma)} d\gamma \varphi d(\gamma b - T_{WT}) d\gamma, \quad (9)$$

$$= \tau l(\gamma b) + \int_{\gamma b}^{\gamma b + T_{WT}} e^{P(\gamma b + b - \gamma)} d\gamma \varphi d(\gamma b$$

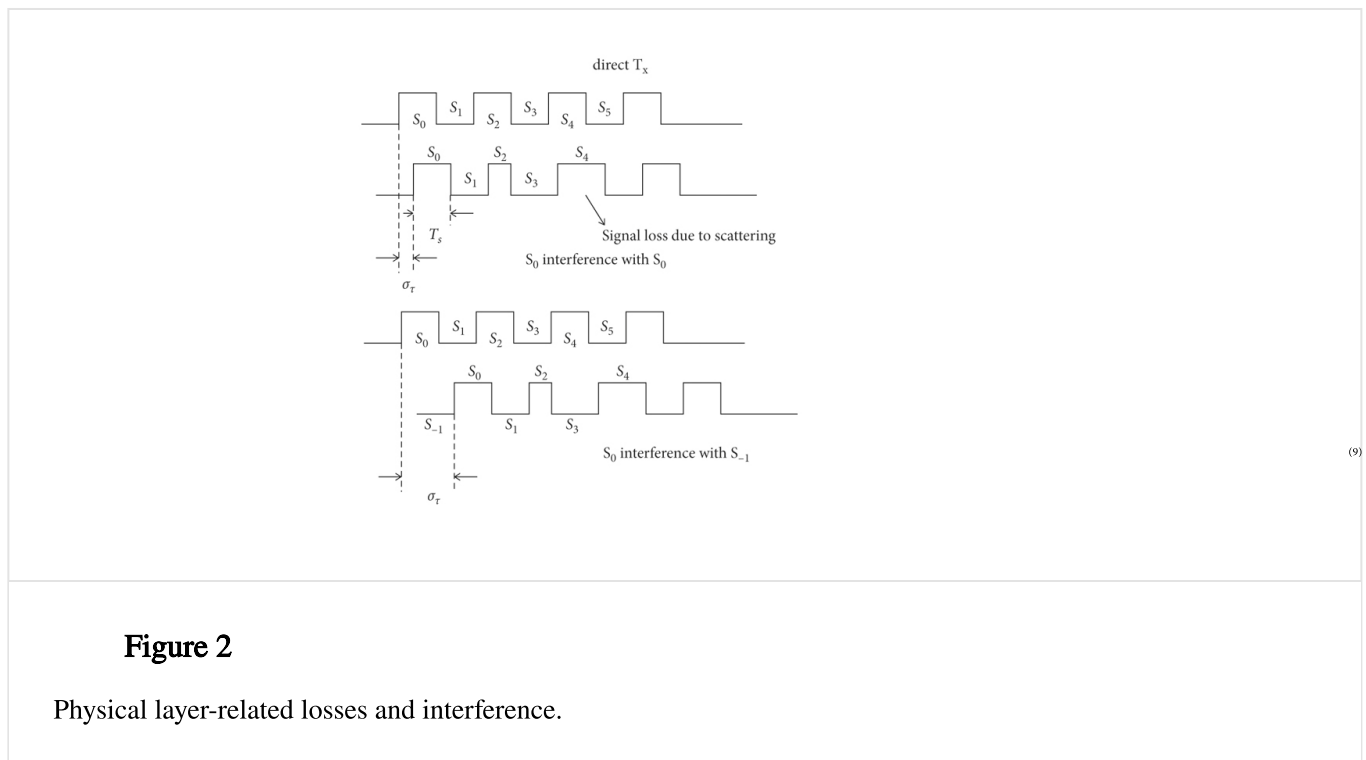
$$- b) + \int_{\gamma b + T_{WT}}^{\gamma b + b} e^{P(\gamma b + b - \gamma)} d\gamma \varphi d(\gamma b),$$

$$\gamma^1 = (\gamma b + b - \gamma),$$

$$\tau l(\gamma b) - \int_b^{\gamma b + T_{WT}} e^{P\gamma^1} d\gamma^1 \varphi d(\gamma b - b)$$

In other scenarios, multiple copies of the signal are transmitted in the time interval and they take different routes while travelling (direct path and scattered path). The spread in the delay indicated as σ_τ of 1 to 3 $h(\tau)$ is the delay profile. Taking the Fourier transform, we obtain $H(f) = \int_0^\infty h(\tau)e^{-j2\pi f\tau} d\tau$. The coherence bandwidth (at which the delay profile response is almost flat, if the signal bandwidth $\beta_s < \beta_c$ is less than coherence of $\beta_c = 1/2\sigma_\tau$.

Figure 2 shows the signal for $\sigma_\tau \ll T_s$ or $\sigma_\tau \gg T_{\text{signals}}$. So, the sound signal interferes each other significantly and so on as delay spread increases to $\sigma_\tau > T_{\text{signals}}$ and $1/T_{\text{signals}} > 1/\sigma_\tau$, which implies to $\beta_s > 2\beta_c$ obtained as the interference. The T_x is stable and R_x is moving towards T_x , indicating the change in the frequency of sound varying due to relative motion between the T_x and R_x .



5.2. Simulation Results

In this section, we discuss the performance comparison of the proposed EELORP protocol by conducting simulations in Aqua-Sim [63–66]. Aqua-Sim is an extended version of NS-2 and offers easy implementation of underwater network scenarios. The parameters used for setting up the network are given in Table 2.

Table 2

Using the simulations, we measure the energy consumption in nodes and the delay that occurred in the transmission of data in the UWSN. We also compare the results obtained by our proposed work with vector-based forwarding (VBF). Figure 3 shows the energy consumption by nodes in the network. From the results obtained, we can see that the nodes consume less energy using the proposed EELORP protocol compared to VBF protocol. Initially, the nodes have the same level of energy consumption with both the protocols, but as the number of nodes increases, the energy consumption using VBF becomes more compared to the proposed scheme. This signifies the better energy efficiency offered by the proposed protocol.

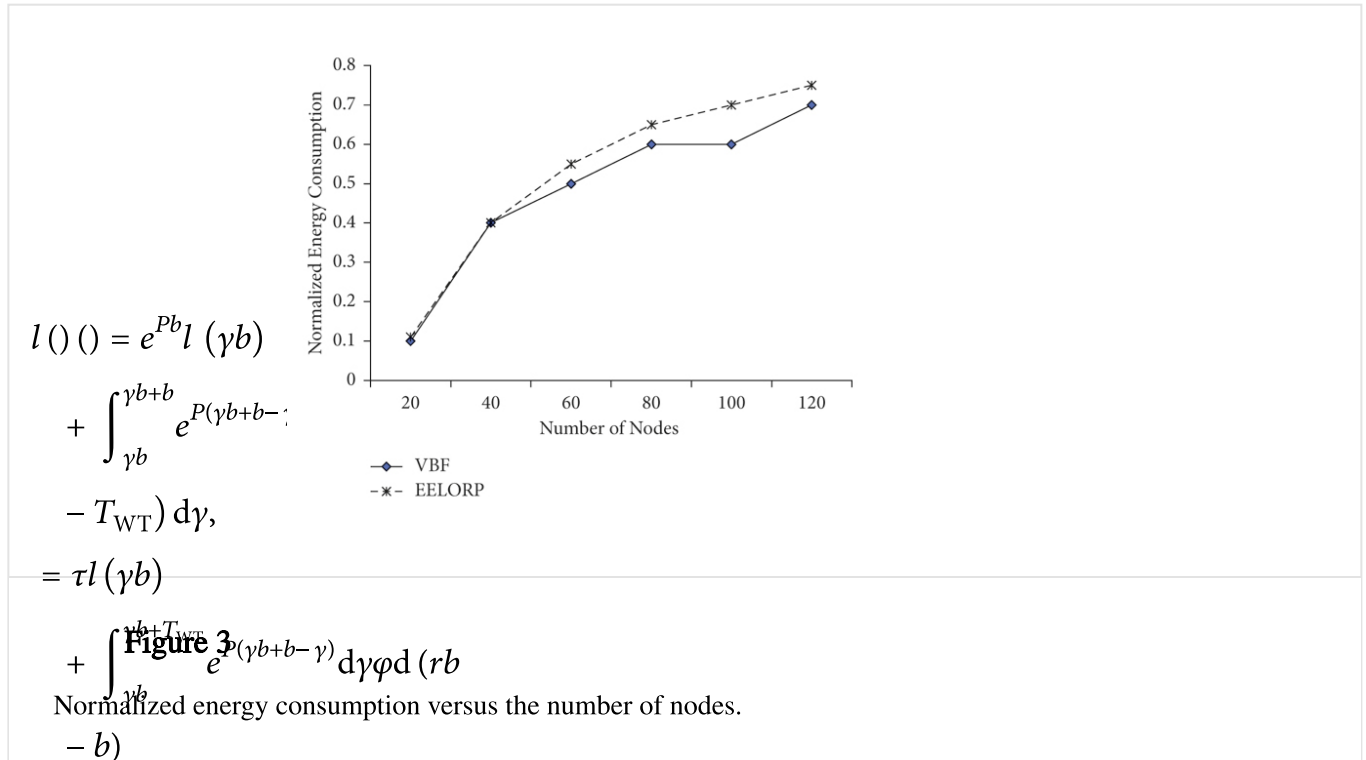


Figure 4 shows the delay incurred in transmission of the data packets using the protocols in the UWSN. From the results, we can see that using the proposed method EELORP and VBF, the delay incurred remains almost similar when the number of nodes is less. But as the number of nodes increases, the EELORP has less delay compared to VBF in the network. Thus, our results show that the proposed protocol can be used efficiently for numerous possibilities in underwater acoustic sensor networks with reduced delay.

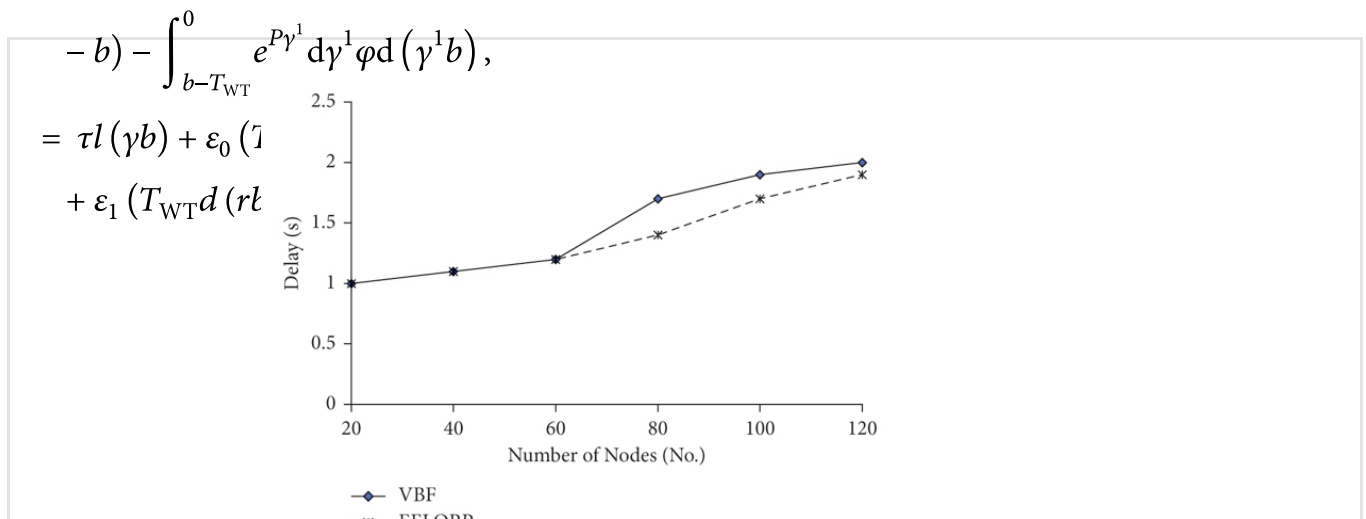


Figure 4

Delay versus number of nodes.

6. Future Research Directions

6.1. Energy Efficiency

This has emerged as one of the major research areas for opportunistic routing protocols in underwater acoustic sensor networks. With restrictions and various limitations in recharging the sensor nodes, it is very important for any routing protocol to minimize the energy usage in the nodes while ensuring that the data gets delivered to the destination. Numerous protocols have tried to improve the energy efficiency in the network, but as UASN has an unpredictable nature, we should for further improvement in this research direction.

6.2. Channel Utilization

The unique features of UASNs like high propagation delay, constant mobility of sensor nodes, high error rate, and interference lead to a major challenge in ensuring the efficient utilization of the channel. Most of the existing protocols have various limitations in channel utilization and this area would be a major area of focus.

6.3. Communication Holes

Dealing with communication holes is a major challenge in UASNs, especially in networks with sparse deployment. Frequent movement of the sensor nodes due to currents and other reasons and failure of sensor nodes due to energy drainage or damage create void areas in the network. Thus, nodes will be unable to find suitable neighbor nodes to forward the data packet to the destination.

6.4. Security

Security of the data transmitted has become one of the major requirements of the applications deploying UASNs. It is therefore vital for all the routing protocols to include a security mechanism that can secure the data from any attackers.

6.5. Reliable Delivery

Reliable delivery of data packet at the destination is a major challenge in UASNs with a dynamic environment. Due to multiple reasons like damage of nodes, lack of energy, voids, etc., the data packet might get lost in the network. It is very important for any protocol to have strategies to manage any data loss and to make sure that the data reaches the destination, also keeping the number of retransmissions to a minimum to save the energy of nodes.

This study presented a systematic survey on the location-based opportunistic routing protocols in underwater acoustic sensor networks. The study initially discussed the working of underwater sensor networks, the challenges and issues, the latest applications using UASNs, and the working of opportunistic routing in underwater acoustic sensor networks. A detailed discussion on all recently proposed location-based opportunistic routing protocols was presented with a focus on their design and working. A discussion on the design and working of an easy-to-implement energy-efficient location-based opportunistic routing protocol (EELORP) that can be used efficiently for numerous possibilities in underwater acoustic sensor networks with reduced delay was presented. A discussion on results obtained with simulations was then presented along with comparisons with existing protocols. Finally, a brief discussion on the future research directions was presented.

Data Availability

All the output data with details can be accessed through the first author at reasonable request (email: varunmenon@ieee.org).

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. T. Qiu, Z. Zhao, T. Zhang, C. Chen, and C. L. P. Chen, "Underwater Internet of things in smart ocean: system Architecture and open issues," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4297–4307, 2000.
View at: [Google Scholar](#)
2. G. Han, C. Zhang, L. Shu, and J. J. P. C. Rodrigues, "Impacts of deployment strategies on localization performance in underwater acoustic sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1725–1733, 2015.
View at: [Publisher Site](#) | [Google Scholar](#)
3. S. Fattah, A. Gani, I. Ahmedy, M. Y. I. Idris, and I. A. Targio Hashem, "A survey on underwater wireless sensor networks: requirements, taxonomy, recent advances, and open research challenges," *Sensors*, vol. 20, no. 18, p. 5393, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
4. G. Han, X. Long, C. Zhu, M. Guizani, and W. Zhang, "A high-availability data collection scheme based on multi-AUVs for underwater sensor networks," *IEEE Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1010–1022, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
5. A. M. Khasawneh, O. Kaiwartya, A. Khalifeh, L. M. Abualigah, and J. Lloret, "Green computing in underwater wireless sensor networks pressure centric energy modeling," *IEEE Systems Journal*,

6. V. Menon, D. Midhunchakkaravarthy, S. Verma, A. Sujith, and M. Manju, "Enabling reliable communication in Internet of underwater things: applications, challenges and future directions," in *Proceedings of the 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*, pp. 296–301, Jalandhar, India, May 2021.
View at: [Publisher Site](#) | [Google Scholar](#)
7. G. Han, H. Wang, J. A. Ansere, J. Jiang, and Y. Peng, "SSLP: a stratification-based source location privacy scheme in underwater acoustic sensor networks," *IEEE Network*, vol. 34, no. 4, pp. 188–195, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
8. V. G. Menon, "Opportunistic routing protocols in underwater acoustic sensor networks: issues, challenges, and future directions," *Magnetic Communications*, CRC Press, FL, USA, pp. 109–130, 2018.
View at: [Publisher Site](#) | [Google Scholar](#)
9. H. Cui, Y. Zhang, X. Liu, and D. Sun, "The simulation and emulation platforms of underwater acoustic sensor networks," in *Proceedings of the IEEE/OES China Ocean Acoustics (COA)*, pp. 1–5, Harbin, China, January 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
10. Y. Chen, X. Jin, and X. Xu, "Mobile data collection paths for node cooperative underwater acoustic sensor networks," in *Proceedings of the OCEANS*, pp. 1–5, Shanghai, China, April 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
11. M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "A survey of architectures and localization techniques for underwater acoustic sensor networks," *IEEE Communications Surveys & Tutorials*, in *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 487–502, 2011.
View at: [Publisher Site](#) | [Google Scholar](#)
12. K. S. Keerthi, B. Mahapatra, and V. G. Menon, "Into the world of underwater swarm robotics: architecture, communication, applications and challenges," *Recent Advances in Computer Science and Communications*, vol. 13, no. 2, pp. 110–119, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
13. J. Yan, Y. Gong, C. Chen, X. Luo, and X. Guan, "AUV-aided localization for Internet of underwater things: a reinforcement-learning-based method," *IEEE Internet of Things Journal* *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9728–9746, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
14. W. Zhang, G. Han, X. Wang, M. Guizani, K. Fan, and L. Shu, "A node location algorithm based on node movement prediction in underwater acoustic sensor networks," *IEEE Transactions on Vehicular Technology*, *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3166–3178, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)

Electronics, vol. 68, no. 2, pp. 1707–1716, 2021.

View at: [Publisher Site](#) | [Google Scholar](#)

16. S. Zheng, X. Cao, F. Tong, G. Zhang, and Y. Dong, “Performance evaluation of acoustic network for underwater autonomous vehicle in confined spaces,” in *Proceedings of the IEEE 8th International Conference on Underwater System Technology: Theory and Applications (USYS)*, pp. 1–4, Wuhan, China, December 2018.

View at: [Publisher Site](#) | [Google Scholar](#)

17. S. Y. Kulik, A. Y. Rodionov, F. S. Dubrovin, and P. P. Unru, “On reliability of data transmission and distance estimation using mobile underwater acoustic modems,” in *Proceedings of the 25th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS)*, pp. 1–4, St. Petersburg, Russia, May 2018.

View at: [Publisher Site](#) | [Google Scholar](#)

18. S. Biswas and R. Morris, “ExOR, ACM SIGCOMM - Computer Communication Review,” in *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 133–144, USA, October 2005.

View at: [Publisher Site](#) | [Google Scholar](#)

19. M. Ismail, M. Islam, I. Ahmad, F. Aslam, and A. Baseer, “Reliable path selection and opportunistic routing protocol for underwater wireless sensor networks,” *IEEE Access*, vol. 8, pp. 100346–100364, 2020.

View at: [Publisher Site](#) | [Google Scholar](#)

20. V. G. Menon and P. M. J. Prathap, “Comparative analysis of opportunistic routing protocols for underwater acoustic sensor networks,” in *Proceedings of the International Conference on Emerging Technological Trends (ICETT)*, pp. 1–5, Kollam, India, October 2016.

View at: [Publisher Site](#) | [Google Scholar](#)

21. P. Xie, J. Cui, and L. Lao, “VBF: vector-based forwarding protocol for underwater sensor networks,” *NETWORKING 2006. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, vol. 3976, pp. 1216–1221, 2006.

View at: [Publisher Site](#) | [Google Scholar](#)

22. H. Daeyoung and D. Kim, “DFR: directional flooding-based routing protocol for underwater sensor networks,” in *Proceedings of the IEEE Oceans*, Quebec, QC, Canada, September 2007.

View at: [Publisher Site](#) | [Google Scholar](#)

23. W. Liang, H. Yu, L. Liu, B. Li, and C. Che, “Information-carrying based routing protocol for underwater acoustic sensor network,” in *Proceedings of the International Conference on Mechatronics and Automation (ICMA 2007)*, pp. 729–734, Harbin, China, August 2007.

View at: [Publisher Site](#) | [Google Scholar](#)

24. N. Nicolaou, A. See, P. Xie, J. H Cui, and D. Maggiorini, “Improving the robustness of location-

View at: [Google Scholar](#)

25. J. Chen, X. Wu, and G. Rebar Chen, "A reliable and energy balanced routing algorithm for UWSNs," in *Proceedings of the 7th International Conference on Grid and Cooperative Computing (GCC'08)*, pp. 349–355, Shenzhen, China, October 2008.

View at: [Google Scholar](#)

26. P. Xie, Z. Zhou, Z. Peng, J. -H. Cui, and Z. Chi, "Void Avoidance in three-dimensional mobile underwater sensor networks," in *In Proceedings of the the 4th International Conference on Wireless Algorithms, Systems and Applications*, Boston, MA, USA, August 2009.

View at: [Publisher Site](#) | [Google Scholar](#)

27. M. V. PriyaA. A. Kumari, "Traffic aware multipath communication for time-critical applications in underwater acoustic sensor networks," *International Journal of Managment, IT and Engineering*, vol. 2, pp. 66–73, 2012.

View at: [Google Scholar](#)

28. P. Wang, D.-hao Fu, C.-qing Zhao, J.-chun Xing, Qi-liang Yang, and X.-fei Du, "A reliable and efficient routing protocol for Underwater Acoustic Sensor Networks," in *Proceedings of the IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, Nanjing, China, May 2013.

View at: [Publisher Site](#) | [Google Scholar](#)

29. J. She, J. Wang, J. Wang, J. Zhang, and S. Wang, "Location-aware routing protocol for underwater sensor networks," in *Proceedings of the Advanced Technologies, Embedded and Multimedia for Human-Centric Computing*, Springer, New York, NY, USA, November 2014.

View at: [Publisher Site](#) | [Google Scholar](#)

30. J. Park, S. Lee, D. Kim, and Y. Hong, "QoS-aware directional flooding-based routing for underwater wireless sensor networks," in *Proceedings of the WUWNET 14th International Conference on Underwater Networks and Systems*, Rome, Italy, November 2014.

View at: [Publisher Site](#) | [Google Scholar](#)

31. I. Naveed, A. TurkiAli, N. F. Muhammad et al., "SEDG: Scalable and Efficient Data Gathering Routing Protocol for Underwater WSNs," *Procedia Computer Science*, Elsevier, Amsterdam, Netherland, vol. 2, pp. 568–575, 2015.

View at: [Google Scholar](#)

32. I. Naveed, A. Turki, N. F. Muhammad et al., "AEDG: AUV-aided Efficient Data Gathering Routing Protocol for Underwater Wireless Sensor Network," *Elsevier, Procedia Computer Science*, vol. 2, pp. 568–575, 2015.

View at: [Google Scholar](#)

33. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Modeling and analysis of opportunistic routing in low duty-cycle underwater sensor networks," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile*

34. S. Basagni, C. Petrioli, R. Petroccia, and D. S. Carp, "A channel-aware routing protocol for underwater acoustic wireless networks," *Ad Hoc Networks*, vol. 34, pp. 92–104, 2015.
View at: [Publisher Site](#) | [Google Scholar](#)
35. Q. Wang, C. Fei, L. Zhi, and Q. Qian, "A novel efficient forwarding protocol for 3-D underwater wireless sensor networks," in *In Proceedings of the IEEE 11th International Conference on Industrial Electronics and Applications*, Hefei, China, June 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
36. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Geographic and opportunistic routing for under water sensor networks," *IEEE Transactions on Computers*, vol. 65, pp. 548–561, 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
37. D. Li, J. Du, and L. Liu, "A data routing algorithm based on Markov model in underwater wireless sensor networks," in *Proceedings of the IEEE 16th International Conference on Ubiquitous Wireless Broadband*, Nanjing, China, December 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
38. S. Han, Y. Yue, W. Meng, and X. Wu, "A localization-based routing protocol for dynamic underwater sensor networks," in *Proceedings of the IEEE Global Communications Conference*, Washington, DC, USA, December 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
39. N. Kanthimathi and Deje, "Void handling using Geo-Opportunistic Routing in underwater wireless sensor networks," *Computers & Electrical Engineering*, vol. 64, pp. 365–379, 2017.
View at: [Publisher Site](#) | [Google Scholar](#)
40. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "Performance modeling and analysis of void-handling methodologies in underwater wireless sensor networks," *Computer Networks*, vol. 126, pp. 1–14, 2017.
View at: [Publisher Site](#) | [Google Scholar](#)
41. M. Yusuf and S. Uddin, "Low-overhead range-based 3D localization technique for underwater sensor networks," in *Proceedings of the IEEE International Conference on Communications*, Kuala Lumpur, Malaysia, May 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
42. E. Isufi, H. Dol, and G. Leus, "Advanced flooding-based routing protocols for underwater sensor networks," *EURASIP Journal on Applied Signal Processing*, vol. 2016, pp. 1–12, 2016.
View at: [Publisher Site](#) | [Google Scholar](#)
43. K. Wang, H. Gao, X. Xu, J. Jiag, and D. Yue, "An energy-efficient reliable data transmission scheme for complex environmental monitoring in underwater acoustic sensor networks," *IEEE Sensors Journal*, vol. 16, pp. 4051–4062, 2016.
View at: [Publisher Site](#) | [Google Scholar](#)

Conference on Advanced Information Networking and Applications, Taipei, Taiwan, May 2017.

View at: [Publisher Site](#) | [Google Scholar](#)

45. I. Yazgi and B. Baykal, "Topology control vector based forwarding algorithm for underwater acoustic networks," in *Proceedings of the IEEE 24th International Conference on Signal Processing and Communication Application*, Zonguldak, Turkey, May 2016.

View at: [Publisher Site](#) | [Google Scholar](#)

46. F. A. Salti, N. Alzeidi, and B. R. Arafeh, "EMGGR: an energy-efficient multipath grid-based geographic routing protocol for underwater wireless sensor networks," *Wireless Networks*, vol. 23, no. 4, pp. 1301–1314, 2017.

View at: [Publisher Site](#) | [Google Scholar](#)

47. N. Kanthimathi and Deje, "Balanced and Multi-objective Optimized Opportunistic Routing for Underwater Sensor Networks," *Wireless Personal Communications*, vol. 94, no. 4, pp. 2417–2440, 2017.

View at: [Publisher Site](#) | [Google Scholar](#)

48. A. Khan, N. Javaid, I. Ali, and H. A. Mohammad, "An energy efficient interference aware routing protocol for underwater WSNs," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 11 10, 2017.

View at: [Google Scholar](#)

49. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Coutinho, "PCR: a power control-based opportunistic routing for underwater sensor networks," in *Proceedings of the MSWiM 2018 - Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 173–180, New York, NY, USA, October 2018.

View at: [Publisher Site](#) | [Google Scholar](#)

50. S. M. Ghoreyshi, A. Shahrabi, and T. Boutaleb, "A stateless opportunistic routing protocol for underwater sensor networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8237351, 18 pages, 2018.

View at: [Publisher Site](#) | [Google Scholar](#)

51. Z. Jin, N. Wang, Y. Su, and Q. Yang, "A glider-assisted link disruption restoration mechanism in underwater acoustic sensor networks," *Sensors*, vol. 18, no. 2, 2018.

View at: [Publisher Site](#) | [Google Scholar](#)

52. W. Zhuo, H. Guangjie, Q. Hongde, Z. Suping, and S. Yancheng, "An energy - aware and void-avoidable routing protocol for underwater sensor networks," *IEEE Access, Journals & Magazines*, vol. 6, pp. 7792–7801, 2018.

View at: [Google Scholar](#)

53. R. Bu, S. Wang, and H. Wang, "Fuzzy logic vector-based forwarding routing protocol for underwater acoustic sensor networks," *Trans. Emerg. Telecommun. Technol.*, vol. 29, pp. 1–18, 2018.

54. F. Ahmed, Z. Wadud, N. Javaid, N. Alrajeh, M. S. Alabed, and U. Qasim, "Mobile sinks assisted geographic and opportunistic routing-based interference avoidance for underwater wireless sensor network," *Sensors*, vol. 18, no. 4, 2018.
View at: [Publisher Site](#) | [Google Scholar](#)
55. Z. Rahman, F. Hashim, M. F. A. Rasid, and M. Othman, "Totally opportunistic routing algorithm (TORA) for underwater wireless sensor network," *PLoS One*, vol. 13, no. 6, 2018.
View at: [Publisher Site](#) | [Google Scholar](#)
56. R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, and A. A. F. Loureiro, "EnOR: Energy balancing routing protocol for underwater sensor networks," in *Proceedings of the IEEE International Conference on Communications*, Paris, France, May 2017.
View at: [Publisher Site](#) | [Google Scholar](#)
57. I. U. Khan, M. Islam, M. Ismail et al., "Adaptive hop-by-hop cone vector-based forwarding protocol for underwater wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 9, Article ID 15501477209, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
58. M. Zhang and W. Cai, "Energy-efficient depth based probabilistic routing within 2-hop neighborhood for underwater sensor networks," in *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1–4, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
59. Y. Chen, J. Zhu, L. Wan, S. Huang, X. Zhang, and X. Xu, "ACOA-AFSA fusion dynamic coded cooperation routing for different scale multi-hop underwater acoustic sensor networks," in *IEEE Access*, vol. 8, Article ID 186788, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
60. Y. Chen, K. Zheng, X. Fang, L. Wan, and X. Xu, "QMCR: a Q-learning-based multi-hop cooperative routing protocol for underwater acoustic sensor networks," in *China Communications*, vol. 18, no. 8, pp. 224–236, 2021.
View at: [Publisher Site](#) | [Google Scholar](#)
61. D. Zhao, G. Lun, and R. Xue, "Coding-aware opportunistic routing for sparse underwater wireless sensor networks," *IEEE Access*, vol. 9, Article ID 50187, 2021.
View at: [Publisher Site](#) | [Google Scholar](#)
62. Z. Jin, Q. Zhao, and Y. Luo, "Routing void prediction and repairing in AUV-assisted underwater acoustic sensor networks," *IEEE Access*, vol. 8, Article ID 54212, 2020.
View at: [Publisher Site](#) | [Google Scholar](#)
63. M. R. Khosravi, H. Basri, and H. Rostami, "Efficient routing for dense UWSNs with high-speed mobile nodes using spherical divisions," *The Journal of Supercomputing*, vol. 74, no. 2, pp. 696–716, 2018.
View at: [Publisher Site](#) | [Google Scholar](#)

Supercomputing, vol. 74, no. 11, pp. 6184–6200, 2018.

View at: [Publisher Site](#) | [Google Scholar](#)

65. M. R. Khosravi, “The shortfalls of underwater sensor network simulators,” *Sea Technology*, vol. 60, no. 5, p. 41, 2019.

View at: [Google Scholar](#)

66. P. Xie, Z. Zhou, Z. Peng et al., “Aqua-Sim: an NS-2 based simulator for underwater sensor networks,” in *Proceedings of the Oceans 2009*, pp. 1–7, Biloxi, MS, USA, October 2009.

View at: [Publisher Site](#) | [Google Scholar](#)

Copyright

Copyright © 2022 Varun G. Menon et al. This is an open access article distributed under the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 Download other formats ✓

 Order printed copies →



Related articles

No related content is available yet for this article.

Follow us:



[Partnerships](#)

[Blog](#)

[Journals](#)

[Article Processing Charges](#)

[Print editions](#)

[Authors](#)

[Editors](#)

[Reviewers](#)

[Partnerships](#)

[Hindawi XML Corpus](#)

[Open Archives Initiative](#)

[Fraud prevention](#)

[Privacy Policy](#)

[Terms of Service](#)

[Responsible Disclosure Policy](#)

[Cookie Policy](#)

[Copyright](#)

[Modern](#)

[slavery statement](#)

[Cookie Preferences](#)

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 গবেষণা পত্রিকা

Chief Editor (Hon.) : Dr. Upen Rabha Hakacham
Editors (Hon.) : Dr. Lalit Chandra Rabha
Dr. Neeva Rani Phukan

মুখ্য সম্পাদক (অবৈতনিক) : ড" উপেন বাভা হাকাচাম
সম্পাদকস্বয়ং (অবৈতনিক) : ড" ললিত চন্দ্ৰ বাভা
ড" নিভা বাণী 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 ILLICIUM 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. 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.

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	*

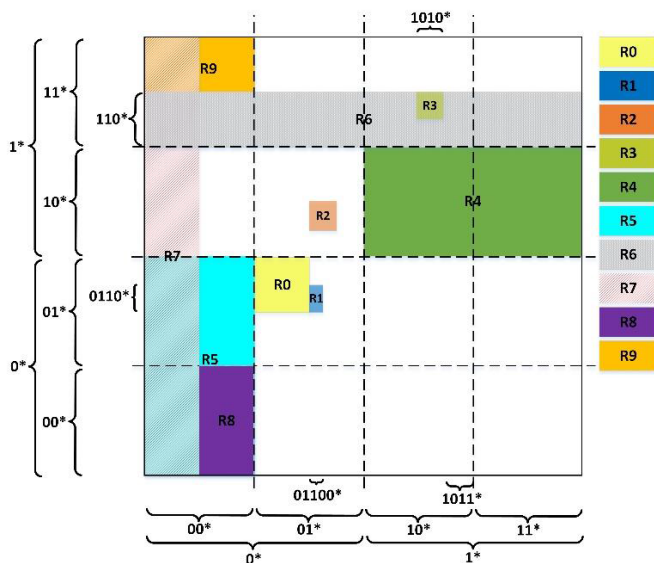


Fig. 1. The rules from Table I as represented in the geometric decomposition space of the KD-tree.

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

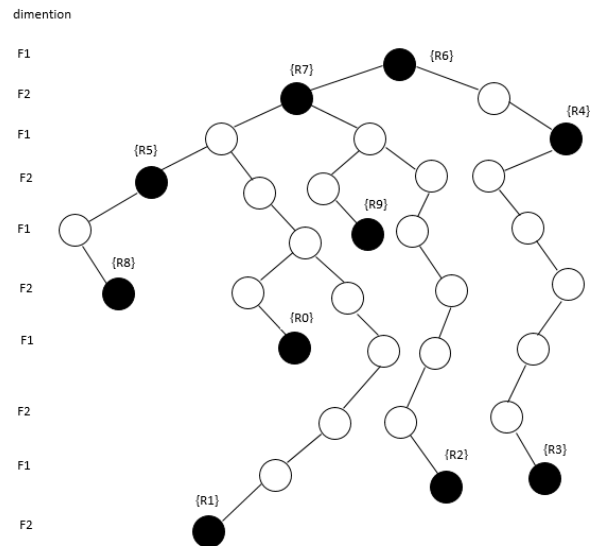


Fig. 2. The binary KD-tree of the geometric space represented in Fig. 1.

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.

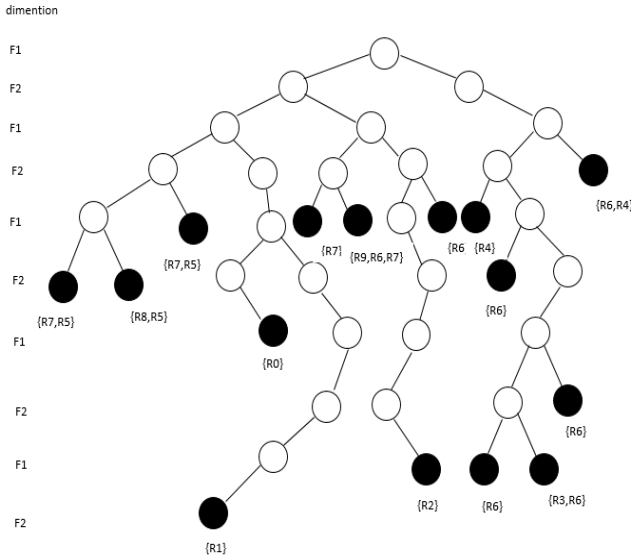


Fig. 3. The leaf-pushed tree of the geometric space represented in Fig. 1.

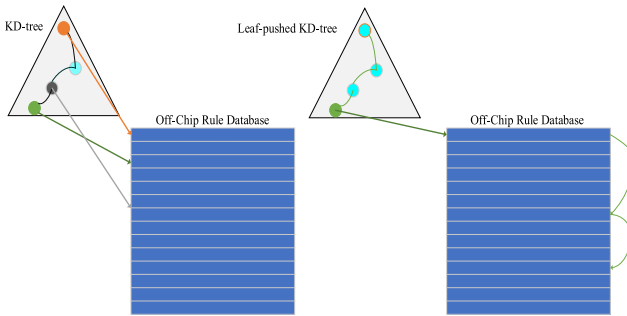


Fig. 4. Comparison of the architecture of conventional KD-tree and leaf-pushing tree.

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

```

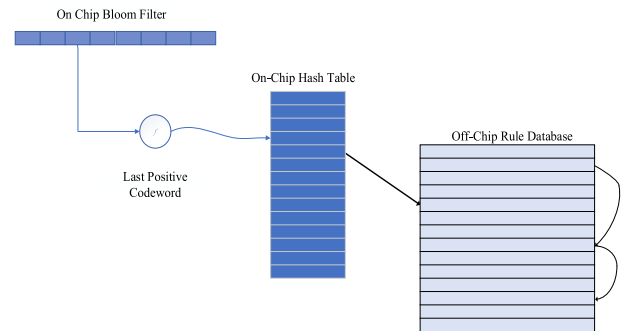


Fig. 5. Searching a tree by means of a Bloom filter and a hash table.

(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

		KD-tree				Leaf-pushed KD-tree			
Size	Rule set Type	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

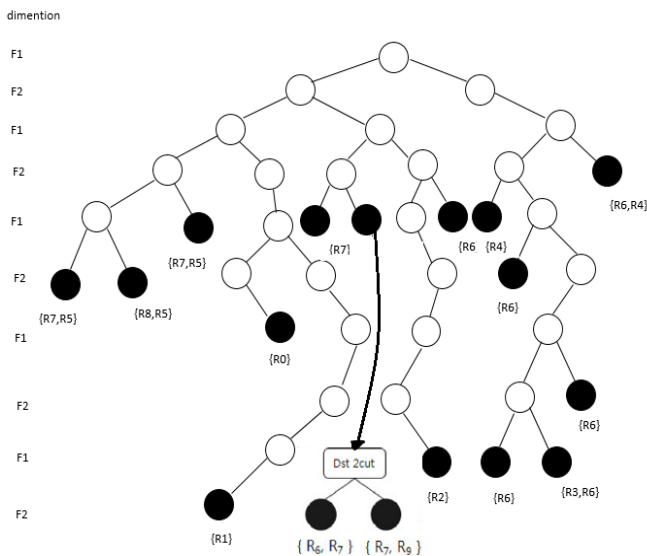


Fig. 6. The structure of the leaf-pushing KD-tree in Figure (3) as modified by means of HiCuts.

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

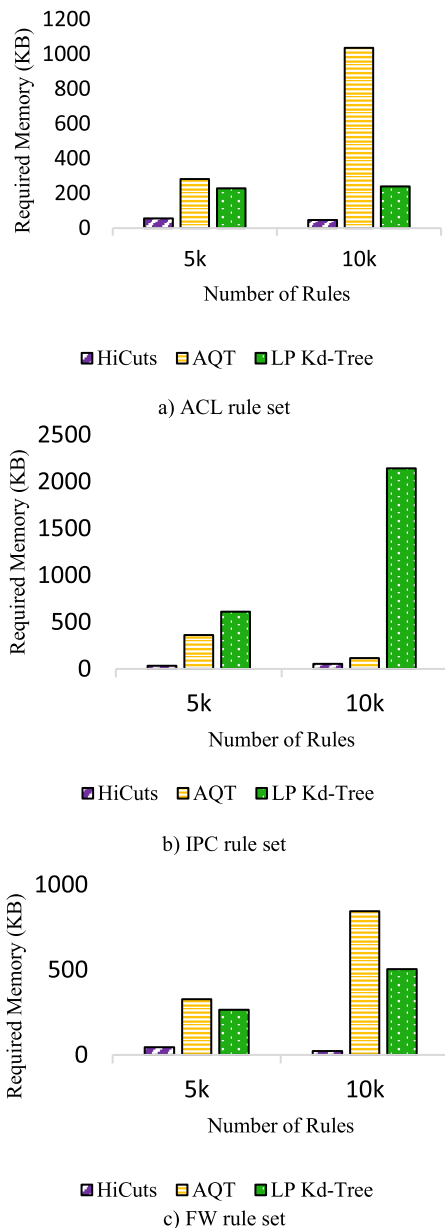


Fig. 7. Comparison of memory access among the proposed algorithm, HiCuts, and AQT.

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 <3, 4, 5, 6, 7, 8, and 9>. 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

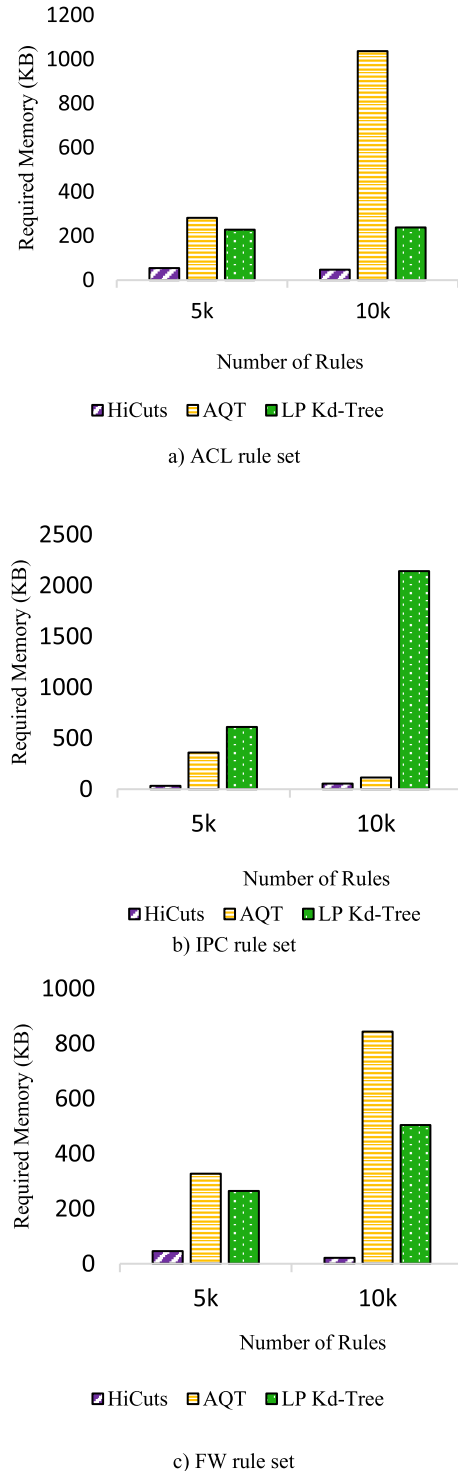


Fig. 8. Comparison of memory usage among the proposed algorithm, HiCuts, and AQT.

(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.

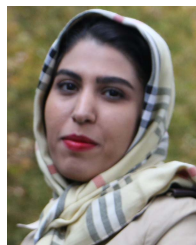
REFERENCES

- [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, "Intelligent 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.



All

ADVA

[Journals & Magazines](#) > [IEEE Consumer Electronics Mag...](#) > [Volume: 11 Issue: 6](#)

Security in Edge-Centric Intelligent Internet of Vehicles: Issues an

Publisher: IEEE

[Cite This](#)

PDF

[Mainak Adhikari](#) ; [Ambigavathi Munusamy](#) ; [Abhishek Hazra](#) ; [Varun G Menon](#) ; [Vijay Anavangot](#) ; [Deepak Puthal](#)

19
Paper
Citations

641
Full
Text Views



[Abstract](#)

Authors

[Figures](#)

[References](#)

[Citations](#)

[Keywords](#)

[Metrics](#)



Downl
PDF

Abstract:Propelled by the growth of automotive industry, and the ubiquity of smart sens systems such as the Internet of Vehicles (IoV) have seen signific... **View more**

► Metadata


Abstract:
Propelled by the growth of automotive industry, and the ubiquity of smart sensors, intellig the Internet of Vehicles (IoV) have seen significant research interest in recent times. The networks support real-time vehicular applications through on-device computing, commun processing, edge computing, and cloud aggregation. While enriching the user experienc

latency through efficient energy management, IoV deployments face the fundamental challenge of security. In this article, we discuss various security attack modes in an edge-centric intelligent IoV framework for distributed smart vehicles, and remote processing units. We highlight various attack detection and mitigation mechanisms for the proposed IoV framework, to address the security challenges. Finally, we discuss research directions to ensure security of sensory data in edge-centric IoV systems.

Published in: IEEE Consumer Electronics Magazine (Volume: 11 , Issue: 6, 01 November 2021)

Page(s): 24 - 31

INSPEC Accession Number: 211000

Date of Publication: 30 September 2021 

DOI: 10.1109/MCE.2021.3110000

ISSN Information:

Publisher: IEEE

Mainak Adhikari

Indian Institute of Information Technology Lucknow, India

Ambigavathi Munusamy

Indian Institute of Information Technology Una, India

Abhishek Hazra

Indian Institute of Technology (ISM) Dhanbad, India

Varun G Menon

SCMS School of Engineering and Technology, India

Vijay Anavangot

Indian Institute of Technology Bombay, India

Deepak Puthal

Newcastle University, UK

Contents

The Internet of Things (IoT) facilitates the omnipresent sharing of information and knowledge between connected devices with limited human interference, which is a crucial instigator for different applications like green infrastructure, smart transportation networks, etc. With this paradigm, the term Internet of Vehicles (IoV) is often encountered in the smart transportation system, referring to the vehicular subsystems that use advanced communication technology, intelligent devices, communication technologies (e.g., 5G and 6G) and intelligent vehicles, sophisticated business analytics, and human-machine cooperation to improve efficiency, performance, and reliability.

[Sign in to Continue Reading](#)

Authors

Mainak Adhikari
Indian Institute of Information Technology Lucknow, India

Ambigavathi Munusamy
Indian Institute of Information Technology Una, India

Abhishek Hazra
Indian Institute of Technology (ISM) Dhanbad, India

Varun G Menon
SCMS School of Engineering and Technology, India

Vijay Anavangot
Indian Institute of Technology Bombay, India

Deepak Puthal
Newcastle University, UK

Figures

References

Citations

Keywords

Metrics

More Like This

Novel design of intelligent internet-of-vehicles management system based on cloud-computing and Internet-of-Things
Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology
Published: 2011

Intelligent Urban Traffic Management System Based on Cloud Computing and Internet of Things
2012 International Conference on Computer Science and Service System
Published: 2012

[CHANGE USERNAME/PASSWORD](#)

[VIEW PURCHASED DOCUMENTS](#)

[COMMUNICATIONS PREFERENCES](#)

[US & CANADA: +1 800 678 4333](#)

[PROFESSION AND EDUCATION](#)

[WORLDWIDE: +1 732 981 0060](#)

[TECHNICAL INTERESTS](#)

[CONTACT & SUPPORT](#)

[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [IEEE Ethics Reporting](#) | [IEEE Privacy Policy](#)

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for humanity.

© Copyright 2023 IEEE - All rights reserved.

IEEE Account

- » [Change Username/Password](#)
- » [Update Address](#)

Purchase Details

- » [Payment Options](#)
- » [Order History](#)
- » [View Purchased Documents](#)

Profile Information

- » [Communications Preferences](#)
- » [Profession and Education](#)
- » [Technical Interests](#)

Need Help?

- » **US & Canada:** +1 800 678 4333
- » **Worldwide:** +1 732 981 0060
- » [Contact & Support](#)

[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [Sitemap](#) | [Privacy & Opting Out of Cookies](#)

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2023 IEEE - All rights reserved. Use of this web site signifies your agreement to the terms and conditions.

Study on Treatment of Blood from Abattoir using Microbial Fuel Cell (MFC) Technology with Production of Green Energy

Article type: Research Article

Authors: [Sreedharan, Sanju](https://content.iospress.com:443/search?q=author%3A%28%22Sreedharan, Sanju%22%29) (<https://content.iospress.com:443/search?q=author%3A%28%22Sreedharan, Sanju%22%29>)

Affiliations: Department of Civil Engineering, SCMS School of Engineering and Technology, Karukutty, Ernakulam, Kerala - 683576, India, sanjus@scmsgroup.org (<mailto:sanjus@scmsgroup.org>).

Abstract: Zero energy technologies and sustainable energy production are the two major concerns of present day researches. Microbial fuel cells (MFCs) are bioreactors that extract chemical energy stored in organic compounds, into electric potential, through bio-degradation. The core reason for the high strength of effluent generated from slaughterhouses is animal blood. The current study evaluates the potential of MFC technology to reduce the pollution strength of cattle blood in terms of chemical oxygen demand (COD). The current study was piloted in three stages using lab scale two chambered MFC: The first stage was to determine the best oxidising agent as compared to natural aeration from three accessible options, KMnO₄, diffused aeration and tape grass aquatic plant. KMnO₄ was found to be the superlative with a 30% reduction in COD in 100 hrs batch reactor and a maximum power of 0.97 mW using 125 mL livestock blood. The second stage of the study optimised the concentration of KMnO₄. At 500 mg/L KMnO₄ concentration, 50% COD removal efficiency was acquired in a batch reactor of 60 hrs with an average energy output of 1.3 mW. In the final stage on the addition of coconut shell activated carbon with an Anolyte at a rate of 40 mL/125 mL of substrate COD removal efficiency increased to 74.9%.

Keywords: Adsorption, bio-energy, cattle blood, microbial fuel cell, wastewater treatment

DOI: 10.3233/AJW210053

Journal: [Asian Journal of Water, Environment and Pollution](https://content.iospress.com:443/journals/asian-journal-of-water-environment-and-pollution) (<https://content.iospress.com:443/journals/asian-journal-of-water-environment-and-pollution>), vol. 18, no. 4, pp. 135-140, 2021

Received 20 May 2020 | 23 March 2021 | **Accepted** 23 March 2021 | **Published:** 18 November 2021

Price: EUR 27.50

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/352993955>

In-Line and Cross-Flow Response Interactions during Vortex Induced Vibration of Marine Risers

Article in MER - Marine Engineers Review · July 2021

CITATIONS

0

READS

96

2 authors:



Vidya Chandran

SCMS School of Engineering and Technology, Karukutty

24 PUBLICATIONS 31 CITATIONS

SEE PROFILE



Sheeja Janardhanan

Indian Maritime University

66 PUBLICATIONS 77 CITATIONS

SEE PROFILE

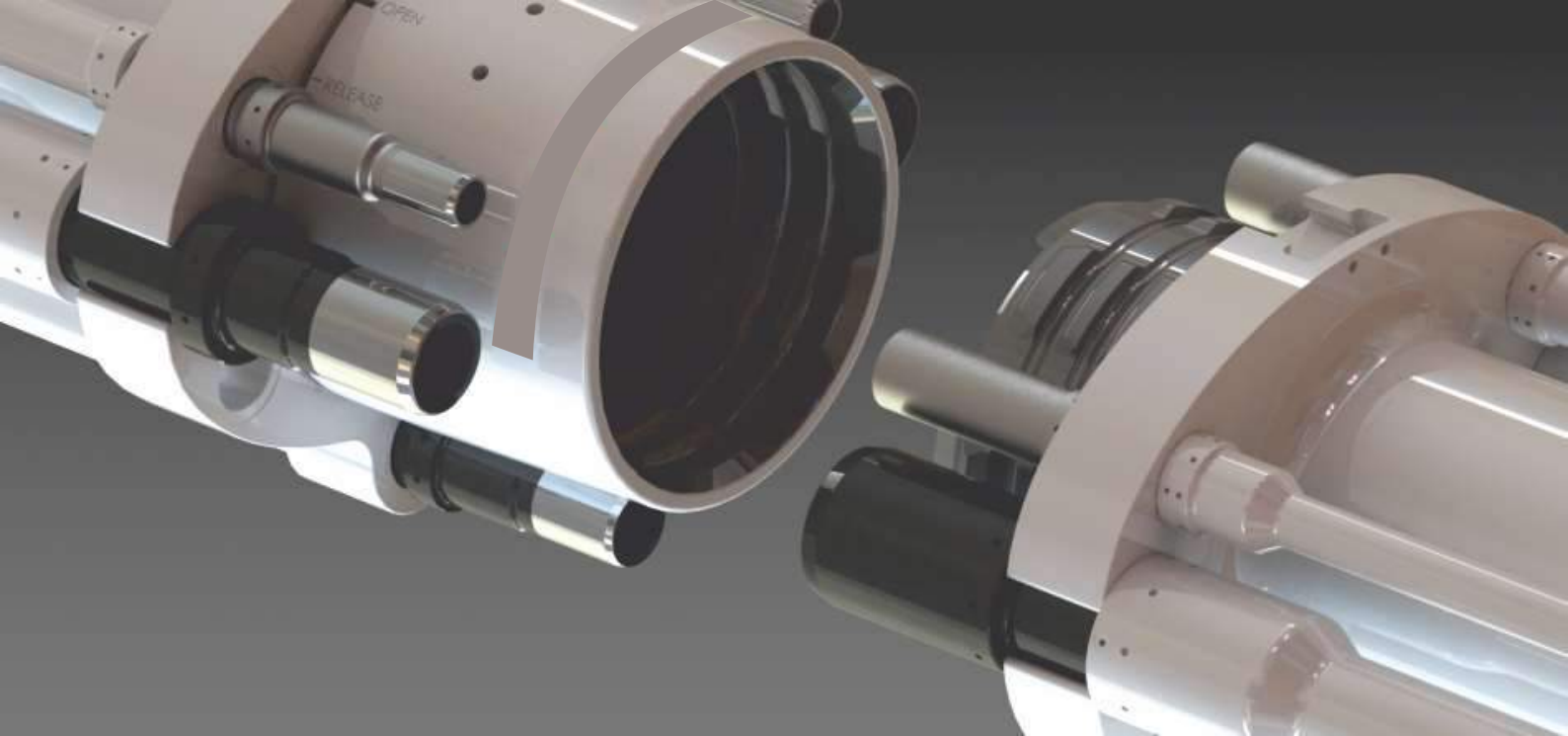
Some of the authors of this publication are also working on these related projects:



Seaplanes [View project](#)



Bio-Inspired Propulsion Systems [View project](#)



In-Line and Cross-Flow Response Interactions during Vortex Induced Vibration of Marine Risers

Vidya Chandran¹, Sheeja Janardhanan²

¹Department of Mechanical Engineering, SCMS School of Engineering and Technology, Karukutty, Kerala, India

²School of Naval Architecture and Ocean Engineering, Indian Maritime University, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

The paper presents a simplified method for understanding the interaction between in-line and cross-flow responses using computational fluid dynamics simulations. Interaction between the responses in the in-line and cross-flow directions in vortex induced vibrations of cylindrical risers in the marine environment is still not fully understood. The trends of variation of hydrodynamic and structural parameters as well as pattern of shedding have been determined numerically to understand the effect of the in-line degree of freedom as well on the riser response and hydrodynamic force coefficients and the results show that a single degree of freedom riser is more susceptible to lock in vibration.

KEYWORDS: Vortex Induced Vibration, In-line, Cross-Flow, Force Coefficients, Response

1. INTRODUCTION

Drilling riser is a pipe laid vertically from the oil well at the ocean bed to the offshore drilling platform. It conveys the drilling fluid and mud to and from the drill site. Marine drilling risers are used especially with floating rigs which are less stable and in particular cases where disconnection of the platform from the seafloor may be required quite often. **Figure 1** shows various layouts of marine risers depending on the constructional specification of platforms. **Figure 2** shows different

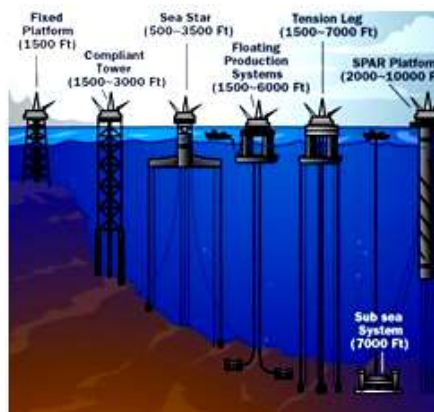


Figure 1 Constructional variation of offshore platforms with drilling depth [1]

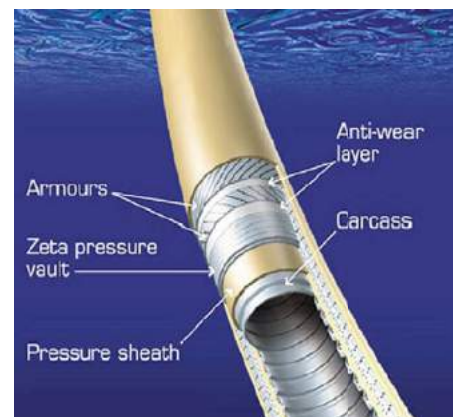


Figure 2 Cross section of a typical flexible riser [1]

Marine drilling risers are used especially with floating rigs which are less stable and in particular cases where disconnection of the platform from the sea floor may be required quite often

layers used in the construction of a flexible riser.

The marine risers, which are classified as **bluff bodies**, and when encountering fluid flow, alternate vortices are shed in the wake of the structure due to boundary layer separation. This alternate vortex shedding happens at a discrete frequency depending upon the flow Reynolds Number (Re). When the vortex shedding frequency matches with the natural frequency of the riser structure, it resonates with high amplitude of oscillation. These large amplitude vibrations, that occurs during "lock-in" of risers are catastrophic and needs to be arrested for the safety of crew working on the floating platforms and also for extending operational life of the risers. Vortex induced vibration (VIV) of marine risers poses all the challenges in the deployment and operation of marine risers.

There have been lot of research in the recent past to understand their behaviour under various sub-sea flow conditions. But most of the studies have concentrated on understanding the wake characteristics and estimating hydrodynamic loading and response of either stationary cylinder or cylinder with a single degree of freedom (1DOF)[2].

Few results have been reported for study of hydrodynamic response of cylinder with two degrees of freedom (2DOF) in both in-line (IL) and cross-flow (CF) directions. IL vibrations have significant impact on the shedding pattern and also on the amplitude of CF vibrations [3].

The first of its kind discussions were reported in the case of flow around cylinder with 2DOF [4]. They established the effect of reduced velocity (U_r) on the effect of forced and free 2DOF response [4]. The effect of IL response on CF response depends on the ratio of natural frequencies in both the directions

$$(\eta_b = \frac{f_{ILCF}}{f_{NL}})$$

During lock in, if the natural frequency in the IL direction is twice that in the CF direction, resonance occurs in both directions leading to premature failure of the riser [5]. Also it has been observed that IL response amplitude is a function of U_r and stability parameter, whereas the CF response amplitude is a function of U_r and flow velocity [6]. Wake characteristics, hydrodynamic force coefficients and response vary significantly when both IL and CF vibrations occur simultaneously. Hence there is a need for prediction of response that hold good for the combined IL and CF vibration.

2. PROBLEM DESCRIPTION

In the present paper a riser model with outer diameter 0.076 m has been numerically analysed using two dimensional (2D) computational fluid dynamics (CFD). Specifications of the riser and the flow condition in listed in **Table 1**. The incoming flow velocity is fixed as 0.5 m/s to maintain the flow regime uniform at $Re = 3.8 \times 10^4$ which corresponds to the ocean condition encountered by a real marine riser used for petroleum extraction in offshore industries [7]. In this paper an effort has been made to study the effect of IL vibration on the amplitude of CF vibration and also on the wake characteristics.

2.1. Mathematical Model

The riser has been modelled as a 2D cylinder with 2DOF in the CF and IL directions. The equations of motion for the riser can be represented as Eq. (1) and (2)

$$m\ddot{Y} + c\dot{Y} + kY = F_L(t) \quad (1)$$

$$m\ddot{X} + c\dot{X} + kX = F_D(t) \quad (2)$$

Where Y is the displacement in CF direction and X is the displacement in

Properties	Values	Units
Diameter (D)	0.076	m
Aspect ratio (L/D)	13.12	-
Flow velocity (V)	0.5	m/s
Reynolds Number of flow (Re)	3.8×10^4	-
Mass ratio (m^*)	0.66	-

Table 1 Riser model specifications and flow characteristics

the IL direction. The excitation forces are lift force, $F_L(t)$ and drag force, $F_D(t)$. The excitation forces are periodic in nature due the alternate shedding of vortices, which causes the riser to oscillate in CF as well as IL directions. The riser is observed to oscillate with frequency equal to frequency of vortex shedding (f_v) in the CF direction and at double the frequency in the IL direction during lock in. Lock in can be defined as the resonance condition during which the vortex shedding frequency lock on to the natural frequency of the riser in the cross flow direction. A simple representation of the mathematical model of riser with 2DOF is represented in **Figure 3**.

The riser is modelled with zero structural damping in the CF and IL directions. k_x and k_y are stiffness coefficients in the IL and CF directions respectively. In the present study $k_x = k_y$. For such a specific case the natural frequencies in both directions will be same and hence $\eta_b = 1$

2.2. Fluid Domain Extends

Figure 4 (a) shows the computational domain for the CFD simulation of VIV of an elastically mounted cylinder with 2DOF. The origin of the Cartesian coordinate system is located at the centre of the cylinder. The length of the

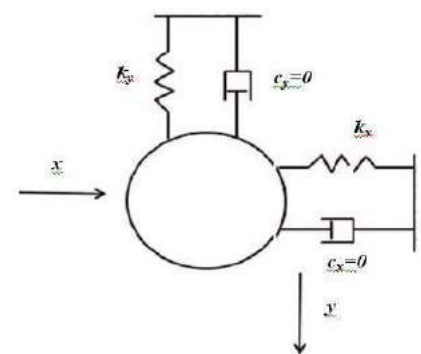


Figure 3 Representation of mathematical model of riser with 2DOF

domain is $40D$ with the cylinder located at $10D$ away from the inlet boundary. The cross flow width of the domain is $20D$ with the centre of the cylinder at the middle.

Detailed views of the mesh around the cylinder along with the computational domain after meshing have been shown in Figure 4 (b) and (c) respectively. There are 307 nodes around the circumference of the cylinder and the minimum element size near the rigid wall boundary has been computed from boundary layer theory to be $0.0001D$ [8].

The non-dimensional element size represented as y^+ , next to the cylinder surface is found to be less than unity. For cylinder wall a no slip boundary condition has been applied assuming the surface to be smooth. Inlet boundary has been treated as velocity-inlet with inflow velocity, $V = 0.5$ m/s. Outlet boundary has been treated as pressure outlet, the gradients of fluid velocity are set to zero and the pressure with zero reference pressure. On the two transverse boundaries, symmetry boundary condition has been applied. Grid independency study has been carried out for the present grid in the previous work done by the authors [9].

2.3. Flow Model

Numerically this problem has been treated as a case of two-way fluid structure interactions (2way FSI). Modeling and meshing has been performed in ANSYS ICEM CFD and solving using ANSYS FLUENT. Flow around the cylinder is modeled using the transient, incompressible Reynolds Averaged Navier-Stokes equation (RANSE) based solver with $k-\omega$ SST as the turbulence model. RANSE solver does the virtual averaging of velocities over an interval of time and hence for a specific interval, the velocity vector appears to be constant in a RANSE solver. In the present work an optimised fine grid is used to compensate for this drawback of the solver enabling it to capture the physics of Von-Karman Street eddies.

The governing equations are discretised using finite difference method. Non iterative time advancement (NITA) scheme with fractional time stepping method (FSM) has been chosen for pressure-velocity coupling of the grid. A least-squares- cell (LSC) based scheme has been used for gradient in spatial discretisation and a second order upwind scheme as convective scheme.

2.4. Structural Model

An elastically mounted cylinder can be mathematically represented by Eq. (1) and (2). These equations of motion are solved using a six degrees of freedom solver (6DOF), an integral part of the main solver by defining the cylinder as an object with 2DOF in transverse direction. A user defined function (UDF) compiled in C programming language has been hooked to the cylinder dynamic boundary conditions. The governing equations for the motion of the centre of gravity of the cylinder in the CF and IL directions are solved in the inertial coordinate system. Velocity in the CF and IL directions are obtained by performing integration on Eq. (3) and (4).

$$\ddot{y} = \frac{1}{m} \sum F_L \quad (3)$$

$$\ddot{x} = \frac{1}{m} \sum F_D \quad (4)$$

Where \ddot{x} and \ddot{y} , are accelerations in the IL and CF direction respectively, m is the mass of the cylinder and F , resultant fluid force acting on the cylinder in the respective direction. Position of the centre of gravity of the cylinder (CG) is updated after solving the equations of motion of a spring mass system represented by Eq. (1) and (2). Mass of the cylinder is given in the UDF as in Eq. (5) and (6).

$$m = m_b + m_a \quad (5)$$

$$m_a = (1 + C_A)m_b \quad (6)$$

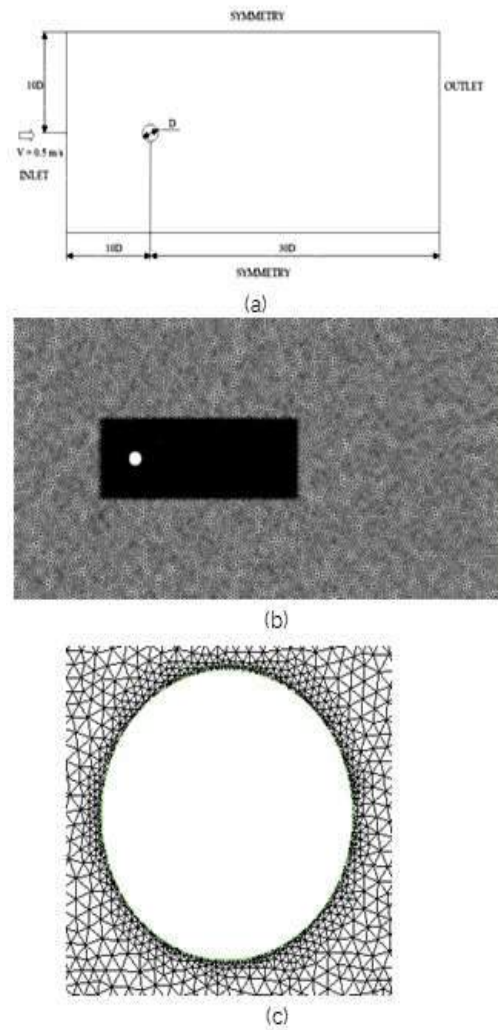


Figure 4 (a) Computational domain
(b) computational mesh
(c) mesh around the cylinder

Where m_a is the added mass and m_b is the mass of the cylinder. Added mass coefficient C_A for the aspect ratio of the present model is found to be equal to 1.0 [10].

Analysis has been performed assigning the cylinder 2DOF with $k_x = k_y$ so that the natural frequencies of the cylinder in both directions remain equal. The results are compared with the case when the cylinder has only 1DOF in the CF direction. Amplitudes of CF response are compared with existing results [9] and also the shedding patterns in both cases have been analysed.

3. RESULTS AND DISCUSSIONS

From the numerical analysis of cylinder with 2DOF it has been observed that the hydrodynamic force coefficient in the CF direction, C_L shows an increase of 17.4% than that for 1DOF case. This result is comparable with the findings of previous research in the field which shows an increase in the lift coefficient value by permitting an extra degree of freedom [11].

RMS value of C_D is constant for both cases with a very small decrease of 4% with 2DOF case. C_L oscillates about zero with almost equal frequencies for both the cases. But the frequency of oscillation of C_D is lesser by 7.2% for 2DOF case. The values of important hydrodynamic and structural parameters of both cases are shown in **Table 2**.

The non-dimensional amplitude in the CF direction obtained with 2DOF is 11.3% more than that with 1DOF. X/D is approximately 0.2. Time histories of major parameters obtained from the 1DOF analysis are shown in **Figures 5(a) – (d)** and that for 2DOF in **Figures 6(a) – (d)**. Frequency of oscillation of the cylinder in the CF direction obtained from 1DOF case is found to be more closer to the theoretical value of vortex shedding frequency obtained from the normal value of $St = 0.2$ ($f_v=1.3$). For the 2DOF case, the frequency of oscillation deviates from the vortex shedding frequency.

For 2DOF case, the frequency of oscillation of C_L and the oscillation frequency of cylinder in the CF direction remains same. In 1DOF case, C_L oscillation frequency remains same as that in the 2DOF case, but the cylinder vibration frequency in the CF shifts towards the natural frequency of cylinder in CF direction.

In the present analysis, the natural frequency in both directions are specifically fixed to be equal to the theoretical value of vortex shedding frequency. Hence the phenomenon can be looked upon as the lock-in of vortex shedding frequency on to the natural frequency of the cylinder. It can be concluded that a cylinder with 1DOF is more prone to lock in vibration compared to that with 2DOF.

This observation can be related to the shifting of the vortex shedding pattern from two singles (2S) to two pairs (2P) mode when motion in IL direction is arrested. The shedding patterns for 1DOF and 2DOF cases are shown in **Figure 5(a)** and **Figure 6(a)** respectively. St obtained also is with the range of normal value for cylinders during lock in. Even though the values of C_D for both cases are almost same, the oscillating frequency varies significantly.

The trajectory of oscillation of cylinder in 2DOF case is represented in **Figure 7**. A clear eight figure trajectory is observed which is typical for VIV of cylinders [11]. Also it has been observed that the motion the IL direction lags behind that in CF direction by a phase angle 30° . The represented trajectory in **Figure 7** corresponds to 30° phase lag [12].

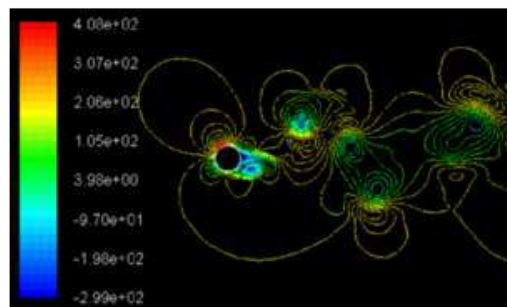
4. CONCLUSIONS

Accounting for an additional degree of freedom seems to have significant effect on the magnitude of lift coefficient but the frequency of oscillation of C_L remains constant for both the cases. C_D is independent of the degree of freedom of the cylinder but the frequency of oscillation varies significantly. Oscillation amplitude of the cylinder in the CF direction is more in 2DOF case which can be related to the increase in C_L .

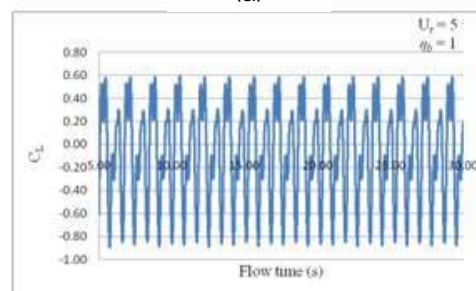
It has been clearly observed that with 1DOF, the cylinder is more susceptible to lock in vibration since the vortex shedding frequency locks on to the natural frequency of the cylinder in the CF direction. But with 2DOF, no such shifting of frequency is observed. Shedding pattern shifts from 2S during 2DOF motion to 2P when motion in IL direction is arrested. An eight figure trajectory typical for VIV is obtained from the

Parameters	1DOF	2DOF ($\eta_b = 1$)
C_L	0.57	0.69
C_D	1.49	1.43
$f_{osc C_L}(f_v)$	1.16	1.14
$f_{osc C_D}$	2.5	2.32
f_{CF}	1.26	1.15
f_{IL}	-	3.28
St	0.18	0.17
Y/D	1.06	1.2
X/D	-	0.17

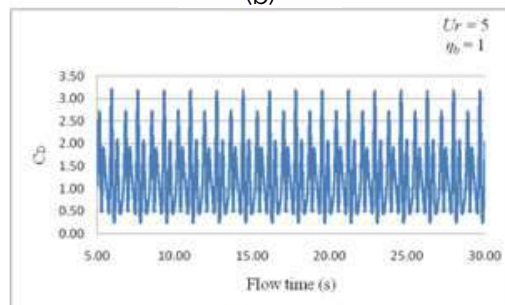
Table 2 Hydrodynamic and structural parameter off cylinder with 1DOF and 2DOF



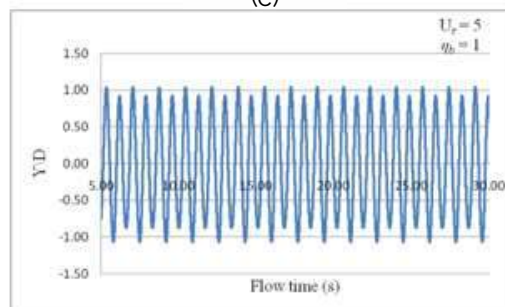
(a)



(b)



(c)



(d)

Figure 5 Pressure contours and Time histories of various hydrodynamic and structural parameters (a) Vortex shedding pattern behind cylinder with 1DOF showing 2P mode (b) C_L of cylinder with 1DOF (c) C_D of cylinder with 1DOF (d) Motion history of cylinder with 1DOF

It can be concluded that a cylinder with 1DOF is more prone to lock in vibration compared to that with 2DOF

2D simulation. Hence the efficacy of 2D CFD as a tool to predict response of cylinder with 2DOF under VIV is accomplished. The observations made above are definitely strong inputs in the design deployment and operation of marine risers.

5. REFERENCES

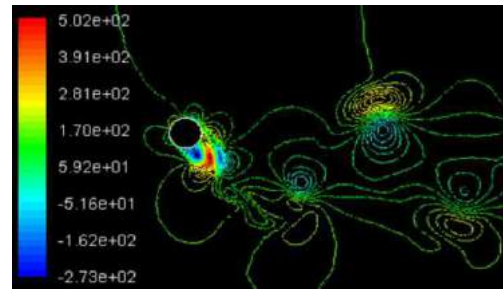
1. Bai, Y., and Qiang B., *Subsea engineering handbook*. Gulf Professional Publishing, Vol. 1, pp. 28, 2018.
2. A. Khalak, C. H. K. Williamson, 'Investigation of the relative effects of mass and damping in vortex induced vibration of a circular cylinder', *Journal of Wind Engineering. Ind. Aerodyn.* vol. 69-71, pp. 341 – 350, 1997.
3. Jauvtis, N. and Williamson, C. H. K., 'The effect of two degrees of freedom on vortex-induced vibration at low mass and damping', *Journal of Fluid Mechanics* 509, 23 – 62, 2004.
4. Moe, G. and Wu, Z. J., 'The lift force on a cylinder vibrating in a current', *Journal of Offshore Mechanics and Arctic Engineering* 112, 297- 303, 1990.
5. T. Sarpkaya., 'Hydrodynamic damping, flow-induced oscillations, and biharmonic response', *ASME Journal of Offshore Mechanics and Arctic Engineering*, 117:232-238, 1995.
6. Yin, D., Experimental and Numerical Analysis of Combined In-line and Cross-flow Vortex Induced Vibrations. *Ph D Thesis, Norwegian University of Science and Technology* pp 7, 2013.
7. Narendran, K., Murali, K., Sundar, V., 'Vortex-induced vibrations of elastically mounted circular cylinder at Re of the O (10^5)', *Journal of Fluids and Structures*, 54, 503 – 521, 2015.
8. Schlichting, H., 'Boundary layer theory', *McGraw-Hill Book Company, New York*, 1979.
9. Chandran, V., Sekar, M., Janardhanan, S., Menon, V., 'A numerical study on the influence of mass and stiffness ratios on the vortex induced motion of an elastically mounted cylinder for harnessing power', *Energies*, 11, 2580, 2018.
10. Naudascher, E.; Rockwell, D., 'Flow induced vibration – An engineering guide', *Dover Publications Inc, Mineola, New York, USA*, 2005.
11. Williamson, C.H.K., Govardhan, R., 'Vortex induced vibrations' *Annual review of fluid mechanics*, vol. 36 pp 413 – 455, 2004.
12. W, Jie., L, Halvor., M, L, Larsen., L, Stergios., B, Rolf. 'Vortex-induced vibration of a flexible cylinder: Interaction of the in-line and cross-flow responses', *Journal of Fluids and Structures* 63 238–258, 2016.

ABOUT THE AUTHORS

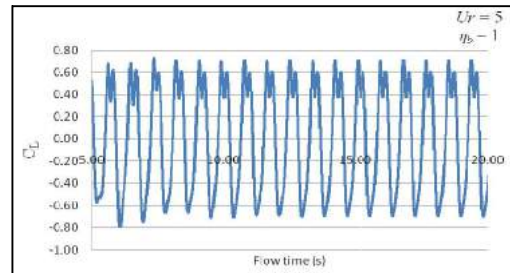
Dr. Vidya Chandran is Associate Professor in the Department of Mechanical Engineering, SCMS School of Engineering and Technology, Cochin. She has a PhD in Vortex Induced Vibrations from Karunya University, India. Her research interests include Vortex Induced Vibrations, Marine Clean Energy and Under Water Robotics.

Dr. Sheeja Janardhanan is Associate Professor in the School of Naval Architecture and Ocean Engineering, Indian Maritime University, Vishakhapatnam, India. She has a PhD in Numerical Ship Hydrodynamics from Indian Institute of Technology Madras, India. Formerly she worked as Professor and Head, Department of Mechanical Engineering, SCMS School of Engineering and Technology, Ernakulam and also as Surveyor in the Research and Rule Development Division of Indian Register of Shipping, Mumbai, India. Her research interests include controllability of surface ships, underwater robotics, vibrations of risers and computational fluid dynamics.

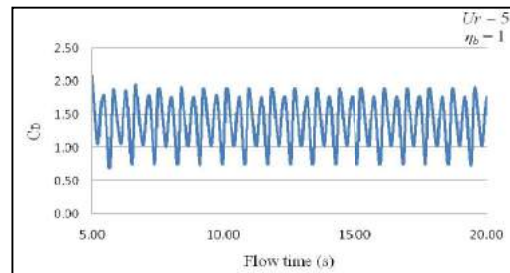
Email: sheeja@imu.ac.in



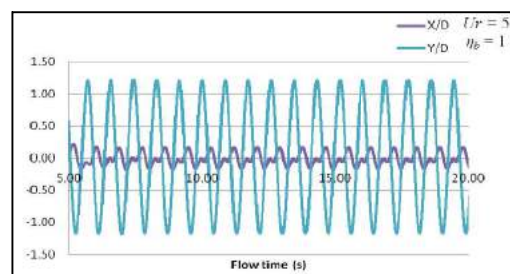
(a)



(b)



(c)



(d)

Figure 6 Pressure contours and time histories of various hydrodynamic and structural parameters (a) Vortex shedding pattern behind cylinder with 2DOF showing 2S mode (b) C_L of cylinder with 2DOF (c) C_D of cylinder with 2DOF (d) Motion history of cylinder with 2DOF

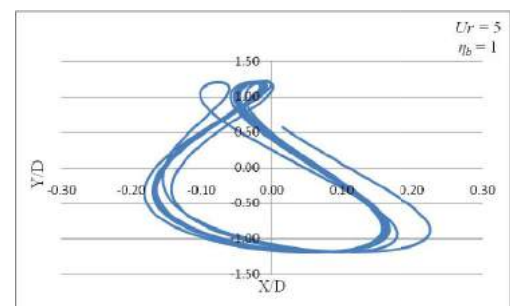


Figure 7 Trajectory of the cylinder with 2DOF motions under VIV