Lecture Notes on Data Engineering
and Communications Technologies 66

A. Pasumpon Pandian
Xavier Fernando
Syed Mohammed Shamsul Islam *Editors*

# Computer Networks, Big Data and IoT

## Proceedings of ICCBI 2020

Springer

# Lecture Notes on Data Engineering and Communications Technologies

Volume 66

The aim of the book series is to present cutting edge engineering approaches to data technologies and communications. It will publish latest advances on the engineering task of building and deploying distributed, scalable and reliable data infrastructures and communication systems.

The series will have a prominent applied focus on data technologies and communications with aim to promote the bridging from fundamental research on data science and networking to data engineering and communications that lead to industry products, business knowledge and standardisation.

Indexed by SCOPUS, INSPEC, EI Compendex.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/15362

A. Pasumpon Pandian · Xavier Fernando ·
Syed Mohammed Shamsul Islam
Editors

# Computer Networks, Big Data and IoT

## Proceedings of ICCBI 2020

Springer

*Editors*
A. Pasumpon Pandian
Department of CSE
KGiSL Institute of Technology
Coimbatore, India

Syed Mohammed Shamsul Islam
Edith Cowan University (ECU)
Joondalup, WA, Australia

Xavier Fernando
Department of Electrical and Computer Engineering
Ryerson University
Toronto, ON, Canada

ISSN 2367-4512 ISSN 2367-4520 (electronic)
Lecture Notes on Data Engineering and Communications Technologies
ISBN 978-981-16-0964-0 ISBN 978-981-16-0965-7 (eBook)
https://doi.org/10.1007/978-981-16-0965-7



This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*We are honored to dedicate the proceedings of ICCBI 2020 to all the participants and editors of ICCBI 2020.*

# Foreword

It is with deep satisfaction that I write this Foreword to the proceedings of ICCBI 2020 held in Vaigai College of Engineering, Madurai, Tamil Nadu, on December 15–16, 2020.

This conference was bringing together researchers, academics and professionals from all over the world, and experts in computer networks, big data and Internet of things.

This conference particularly encouraged the interaction of research students and developing academics with the more established academic community in an informal setting to present and to discuss new and current work. The papers contributed the most recent scientific knowledge known in the field of computer networks, big data and Internet of things. Their contributions helped to make the conference as outstanding as it has been. The local organizing committee members and their helpers put much effort into ensuring the success of the day-to-day operation of the meeting.

We hope that this program will further stimulate research in data communication and computer networks, Internet of things, wireless communication, big data and cloud computing and also provide practitioners with better techniques, algorithms, and tools for deployment. We feel honored and privileged to serve the best recent developments to you through this exciting program.

We thank all the guest editors, authors and participants for their contributions.

Dr. P. Sugumaran
Conference Chair, ICCBI 2020

# Preface

This conference proceedings volume contains the written versions of most of the contributions presented during the conference of ICCBI 2020. The conference provided a setting for discussing recent developments in a wide variety of topics including computer networks, big data and Internet of things. The conference has been a good opportunity for participants coming from various destinations to present and discuss topics in their respective research areas.

This conference tends to collect the latest research results and applications on computer networks, big data and Internet of things. It includes a selection of 74 papers from 248 papers submitted to the conference from universities and industries all over the world. All of accepted papers were subjected to strict peer-reviewing by 2–4 expert referees. The papers have been selected for this volume because of quality and the relevance to the conference.

We would like to express our sincere appreciation to all the authors for their contributions to this book. We would like to extend our thanks to all the referees for their constructive comments on all papers; especially, we would like to thank the organizing committee for their hard work. Finally, we would like to thank the Springer publications for producing this volume.

Coimbatore, India — Dr. A. Pasumpon Pandian
Toronto, Canada — Dr. Xavier Fernando
Joondalup, Australia — Dr. Syed Mohammed Shamsul Islam

# Acknowledgements

ICCBI 2020 would like to acknowledge the excellent work of our conference organizing the committee and keynote speakers for their presentation on December 15–16, 2020. The organizers also wish to acknowledge publicly the valuable services provided by the reviewers.

On behalf of the editors, organizers, authors and readers of this conference, we wish to thank the keynote speakers and the reviewers for their time, hard work and dedication to this conference. The organizers wish to acknowledge Dr. R. Saravanan, Dr. R. Thiruchenthuran, Thiru. S. Kamalakannan, Thiru. S. Balasubramanian and Thiru. S. Singaravelan, for the discussion, suggestion and cooperation to organize the keynote speakers of this conference. The organizers also wish to acknowledge speakers and participants who attend this conference. Many thanks are given to all persons who help and support this conference. ICCBI 2020 would like to acknowledge the contribution made to the organization by its many volunteers. Members contribute their time, energy and knowledge at a local, regional and international level.

We also thank all the chair persons and conference committee members for their support.

# Contents

# About the Editors

**A. Pasumpon Pandian** received his Ph.D. degree in the Faculty of Information and Communication Engineering under Anna University, Chennai, TN, India, in 2013. He received his graduation and postgraduation degree in Computer Science and Engineering from PSG College of Technology, Coimbatore, TN, India, in the year 1993 and 2006, respectively. He is currently working as Professor in the Computer Science and Engineering department of KGiSL Institute of Technology, Coimbatore, TN, India. He has twenty-six years of experience in teaching, research and IT industry. He has published more than 20 research articles in refereed journals. He acted as Conference Chair in IEEE and Springer conferences and Guest Editor in Computers and Electrical Engineering (Elsevier), Soft Computing (Springer) and International Journal of Intelligent Enterprise (Inderscience) Journals. His research interest includes image processing and coding, image fusion, soft computing and swarm intelligence.

**Xavier Fernando** is Professor at the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada. He has (co-)authored over 200 research articles and two books (one translated to Mandarin) and holds few patents and non-disclosure agreements. He was IEEE Communications Society Distinguished Lecturer and delivered close over 50 invited talks and keynote presentations all over the world. He was Member in the IEEE Communications Society (COMSOC) Education Board Working Group on Wireless Communications. He was Chair IEEE Canada Humanitarian Initiatives Committee 2017–2018. He was also Chair of the IEEE Toronto Section and IEEE Canada Central Area. He is a program evaluator for ABET (USA). He was a visiting scholar at the Institute of Advanced Telecommunications (IAT), UK, in 2008, and MAPNET Fellow visiting Aston University, UK, in 2014. Ryerson University nominated him for the Top 25 Canadian Immigrants award in 2012 in which was a finalist. His research interests are in signal processing for optical/wireless communication systems. He mainly focuses on physical and MAC layer issues. He has special interest in underground communications systems, of cognitive radio systems, visible light communications and wireless positioning systems.

**Dr. Syed Mohammed Shamsul Islam** completed his Ph.D. with Distinction in Computer Engineering from the University of Western Australia (UWA) in 2011. He received his M.Sc. in Computer Engineering from King Fahd University of Petroleum and Minerals in 2005 and B.Sc. in Electrical and Electronic Engineering from Islamic Institute of Technology in 2000. Before joining ECU as a Lecturer in Computer Science, he worked in different teaching and research positions at UWA and Curtin University (2011–2016). He was promoted to Senior Lecturer in November 2020. He has published over 60 research articles and got 17 public media releases, including a TV news story and four live radio interviews. He has received the NHMRC Ideas grant 2019 (AUD 467,980) and nine other external research grants. He is serving the scientific community as an Associate Editor of *IEEE Access*, a guest editor of *Healthcare*, a Technical Committee Member of 25 conferences and a regular reviewer of 26 journals. He is a Senior Member of *IEEE* and *Australian Computer Society*. His research interest includes Artificial Intelligence, Computer Vision, Pattern Recognition, Big-Data Analysis, Biometrics, Medical Imaging, Internet of Things (IoT), Image Processing and Biomedical Engineering.

# Maximizing Network Lifetime in WSN Using Ant Colony Algorıthm


**M. D. Saranya, G. Pradeepkumar, J. L. Mazher Iqbal, B. Maruthi Shankar, and K. S. Tamilselvan**

**Abstract** A wireless network is a cluster of specific transducers with statement transportation intend to study it frequently operate in an unpredictable wireless background with vigour constriction. Several types of research are mainly interest in vigour consciousness and statement dependability of a wireless sensor network to maximize network lifetime. In this article, a greedy algorithm and ACO algorithm aims at obtaining the best clarification that satisfies the given set of Greedy algorithm. The aim of the Greedy algorithm obtains a most favorable explanation that satisfies the given set of constraints and also maximizes the given objective function. Ant colony algorithm has been practical to the traveling salesman problem to find the optimal solution in a short time. However, the performance of the ACO algorithm is considered for both high energy efficiency and good power balancing, and maximal energy utilization throughout the network. ACO algorithm it is understandable that the network time increases and extends the life cycle of the wireless sensor network since it manages the energy and power management. The Greedy Algorithm creates the problem of selecting a communication path using the traveling salesman and cracks the logic by using this algorithm. The algorithm uses the single source to all destination technique to find through path for optimum network connectivity. The simulation results were demonstrated with the help of the algorithm and it outperforms the shortest path length concerning the network lifetime.

**Keywords** Wireless network · Maximizing network lifetime · Greedy algorithm · ACO · Communication reliability · Connectivity · Coverage · Traveling salesman problem

M. D. Saranya (✉) · G. Pradeepkumar · K. S. Tamilselvan
Department of ECE, KPR Institute of Engineering and Technology, Coimbatore, Tamilnadu, India

J. L. Mazher Iqbal
Department of ECE, Veltech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India

B. Maruthi Shankar
Department of ECE, Sri Krishna College of Engineering and Technology, Coimbatore, Tamilnadu, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_1

## 1 Introduction

The wireless group is the way of contact along with sensors and observers. A Sensor network is a collection of dedicated transducers among a transportation infrastructure intended to monitor as well as record conditions at the varied location. WSN naturally encompasses numerous integer of spatially detached array-operated, an embedded option that is a network to accumulate the data to the users, and it has the restricted computing and processing capabilities. Full coverage with connectivity suggests that every location in the field is enclosed by at least one node and the information on this position can be running scared to the fusion center. The complexity of data routing and processing also depends on topology. In a WSN the sensors are work in a longer lifetime. Each communication task has an implementation time, relative deadline, and a time. The duration of a WSN is increased by scheduling the vigorous interval of devices. It could affect the performance of the network for connectivity Rc/Rs ratio has to be considered. At each node in the system should form a linked envelop to attain sensing coverage and system connectivity.

## 2 Review of Related Work

General optimization carried at improved force effectiveness and improved network lifetime in wireless sensor network; two critical factors are information packet size and broadcast control level. Conversely, the effect of slighter package size is disintegration into additional information packets and thereby indulgence of increased force [1, 2]. Wireless Sensor Network Lifetime (NL) is a critical metric since the antenna nodes frequently rely on the limited power supply. Cross layer network lifetime maximization consider the joint best possible proposal of the substantial, Medium Access Control (MAC), and system layers to exploit the NL of the power-constrained WSN [4]. The problem of NL maximization can be formulated as a nonlinear optimization problem encompassing the routing stream, link scheduling, contact rate, and control allocation operations for all energetic Time Slots (TSs) [5]. The complexity of NL Maximization (NLM) can be formulated as a mixed integer-convex optimization difficulty with the implementation of the Time Division Multiple Access (TDMA) technique [3, 6]. The wireless sensor system topology, the convenience of resources, and the energy consumption of nodes in different paths of the data collection tree may vary largely, thus affecting the general network lifetime [8, 7]. Cross-layer protocol, which incorporates a multipath routing protocol and an information interleaving practice based on Reed–Solomon code. Formulate the trouble of selecting sensor announcement paths as knapsack trouble and resolve it by a greedy algorithm [10, 9]. The multipath routing protocol then enables all sensors to select multiple statement paths using the proposed optimization algorithm [11, 12]. On the basis of numerous communication paths, the method of data interleaving is working by using a Reed–Solomon code to provide reliable data transmission. Simulation results

can display the available multipath routing protocols to the system lifetime since it balances energy utilization and promotes communication reliability [13, 14].

## 2.1 Routing Protocol

Routing is the method of selecting a path for traffic in several networks. Direction finding is performed for many types of networks, such as the path switched networks, public switched telephone network such as the Internet, as well as in networks used in society and covert transportation, such as the organization of streets, roads, and highways in universal communications. In packet switching networks, routing is the higher-level judgment construction that directs network packets from their source toward their destination through middle network nodes by exact packet forwarding mechanisms.

## 2.2 Traveling Salesman Problem

Traveling Salesman Problem belongs to a set of troubles in computational complexity analysis called NP-complete problem. If could find a way to declare an NP-complete problem and can employ the algorithm to explain all NP problems rapidly. TSP has a number of applications silent in its purest formulation preparation, logistic, and creates of microchips. The challenge of the complexity is that the traveling salesman needs to reduce the total length of the trip. The goal of the Traveling Salesman Problem (TSP) is to find the majority competent way to tour a choose number of "cities". Traveling Salesman Problem conserve be modeled as an undirected weighted graph, such that cities are the graph's vertices. Figure 1 shows the TSP crisis to find the minimum length of the path using the ACO algorithm. The shortest path = a–d–e, Net weight = 22.

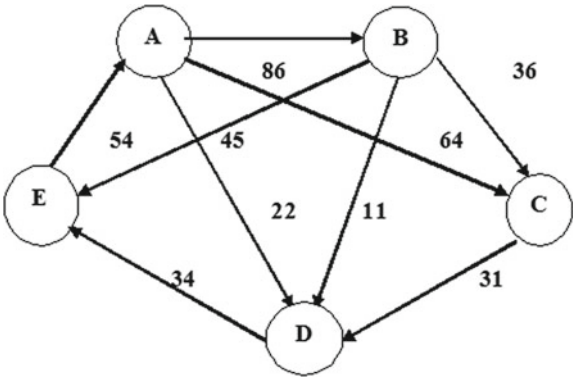



**Fig. 1** Example of TSP problem

# 3 Proposed Scheme

## *3.1 Greedy Algorithm*

In this article, the greedy algorithm aims at obtaining the finest solution that satisfies the given set of constraints and also maximizes the given objective function. A greedy algorithm is a traffic initiator that generates data at the maximum rate feasible and at the earliest chance possible. Each source forever has data to transmit and is not at all in an idle state due to or another local host. Greedy algorithm preserve is characterized as human life from tiny sight and in addition to non-recoverable. A greedy session is a time-limited packet flow or data flow at the maximum possible rate. A greedy source transfer creation simulation model, or a greedy transfer generator, is useful when simulating and analyzing or measuring the lifetime and throughput of a network. These locally optimal solutions will lastly include a globally best resolution. Greedy algorithm is to resolve the problem, it must be that the best explanation to the big difficulty contains the best possible solution to sub-problems.

## *3.2 Ant Colony Algorithm for TSP*

1. ACO is an inhabitants-based Meta-heuristic to facilitate can exist used to locate estimated solutions to the complicated optimization problem.
2. In ACO, a location of a software agent called synthetic ants investigates excellent solutions to a known optimization difficulty. To apply ACO, the optimization trouble is transformed and addicted to the problem of finding the finest path on a weighted diagram.
3. ACO is a probabilistic technique searching for the finest pathway in the graph based on the performance of ants seeking a path connecting their colony and based on food. Ants steer from shell to food source; ants are blind.

Ant Colony Optimization (ACO) studies reproduction systems that obtain stimulation from the performance of authentic ant colonies and which are used to establish discrete optimization problems. The method that ants discover their grub in the straight path is interesting; ants are secreting pheromones to memorize their path. These pheromones disappear with time; whenever an ant finds food, it marks its entrance journey with pheromones.

Pheromones disappear faster on longer paths. Shorter paths make obtainable as the way to food for most of the other ants. The shorter path will be unbreakable by the pheromones further. Finally, the ants revolve up at the shortest path. Ants leave pheromone trail when they make an adaptation trails that are used in prioritizing evolution. The communication along with individuals, or stuck between individuals and the atmosphere, is based on the employ of chemicals produced by the ants. These chemicals are called pheromones.

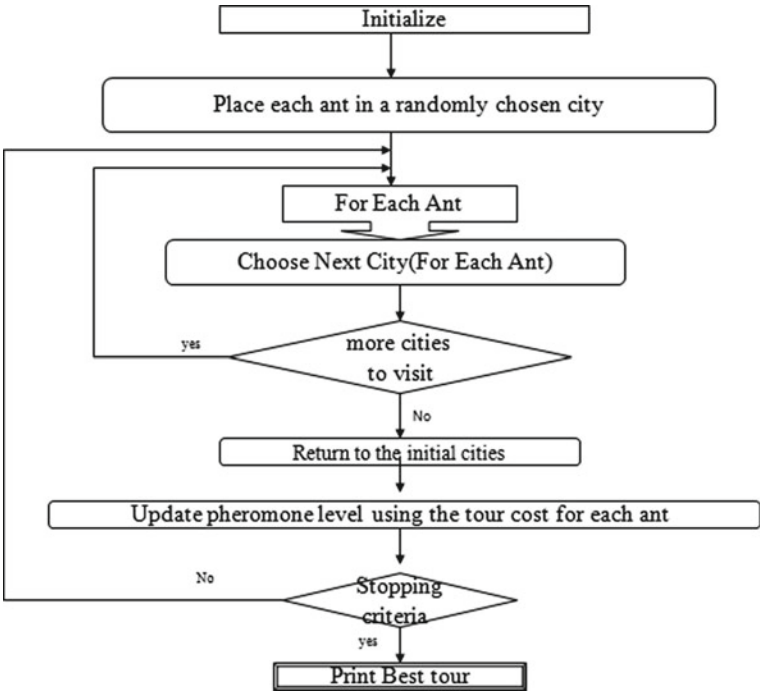



**Fig. 2** Flowchart of traveling salesman predicament

ACO is proficient in solving the Traveling Salesman Problem (TSP). TSP is an NP-hard problem. Figure 2 shows the Flowchart traveling salesman problem given a set of n cities, the Traveling Salesman Problem requires a salesman to discover the shortest way between the given cities and go back to the starting city, while keeping in mind that each metropolitan preserve to be visited only once. The most primitive ACO algorithms used the Traveling Salesman Problem (TSP) as an instance application. The TSP is characteristically represented by a graph $G = (V, E)$, $V$ is the set of integers of vertices, representing the cities, and $E$ being the position of edges that completely connects the vertices. To every edge $(i, j)$ a distance $dij$ is associated. The ACO algorithm, called Ant System (AS), has been applied to the Traveling Salesman Problem (TSP). TSP is also called the Hamiltonian circuit.

## *3.3 Procedure of ACO*

**Step 1**: In the TSP, each ant finds paths in a network and every work is a node. Moreover, no one has to return to the start node and the path is completed when each node is visited.

**Step 2**: Initially, arbitrary levels of pheromone are spread resting on the edges.

**Step 3**: An ant starts at a start node (so the first connection it chooses defines the first task to plan on the machine) as earlier than it uses a conversion regulation to get one step at a time, prejudiced by pheromone levels, and also a heuristic score, each instance choosing the next mechanism to agenda.

**Step 4**: Construct Solutions: Each ant starts at a meticulous state, and then traverse the states one by one.

**Step 5**: Apply Local Search: Before updating the ant's trail, a restricted search can be applied to each solution constructed.

**Step 6**: Update Trails: after the solutions are constructed and calculated, pheromone levels increase and reduce on paths according to favorability.

**Step 7**: Use $d_{ij}$ to indicate the detachment between any two cities in the difficulty. As such $d_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}$.

**Step 8**: Let $\tau_{ij}(t)$ indicate the strength of the track on edge $(i, j)$ at time $t$, at which time every ant will have completed an expedition.

## *3.4 Shortest Path Problem*

Figure 3 shows the shortest path identification to find the minimum length of the path using the ACO algorithm.

The ACO algorithm has been applied to the shortest path problem, and each ant finds paths in a network where each node is referred to as a node. Upon completion of every node, there is no intention to reach the start node once again. Therefore, a total number of available paths in the network and the corresponding path weight get calculated.

(a) **Available Path**

1. Path1—A–B–E–F–I–J
2. Path2—A–B–E–F–G–H–I

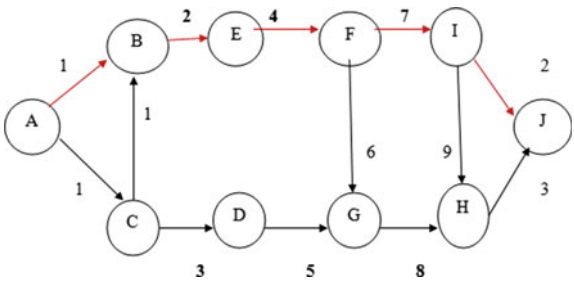



**Fig. 3** Shortest path problem

**Table 1** Number of connected covers

| Number of test case | Number of nodes | Number of connected covers |
|---|---|---|
| Test case 1 | 5 | 4 |
| Test case 2 | 7 | 6 |
| Test case 3 | 10 | 9 |

3. Path3—A–C–B–E–F–I–J
4. Path4—A–C–B–E–F–G–H–J
5. Path5—A–C–D–C–H–J
6. Path6—A–B–E–F–I–H–J
7. Path7—A–C–B–E–F–I–H–J

(b) **Calculate the Path Net Weight**

1. Path1-Netweight-16
2. Path2-Netweight-24
3. Path3-Netweight-26
4. Path4-Netweight-17
5. Path5-Netweight-27
6. Path6-Netweight-25
7. Path7-Netweight-20

(c) **Optimal Shortest Path**

Path1-A-B-E-F-I-J
Netweight-16
No. of. Connected covers-7.

The number of Connected Covers in Shortest Path Identification by Greedy Algorithm is scheduled in Table 1.

## 4 Software Description

### *4.1 Proteus*

Proteus mechanism to test, debug, and regulate a program is problematic. In distinction, greedy simulation allows non-intrusive monitoring and debugging, and also makes it simple to repeat executions. Proteus provides users with unparalleled edibility in choosing or customizing the period of accuracy in the wireless sensor network and memory simulations. Proteus was initially designed for evaluating language, compiler, and runtime system mechanisms to carry portability. Several CAD users release schematics detain as essential immorality in the procedure of creating printed circuit board (PCB) arrangement encloses the disputed in the peak

of investigation. With PCB layout now offering automation of equally section assignment and path routing, getting invent hooked on the computer can frequently be the main occasion overriding constituent of expending even more time working on the schematic. Proteus provides repeatability, non-instructive monitoring and debugging, and incorporated graphical output. The power of its structural design has allowed us to combine first conventional graph-based imitation and now with Proteus VSM—interactive circuit reproduction keen on the invent WSN environment. For the first time eternally, it is sufficient to explain a total circuit for a TSP based Ant Colony Optimization and then test it interactively. It provides in general control of drawing emergence in turns of line widths, fills styles, colours and fonts.

#### 4.1.1 Features of Proteus

1. Regular wire steering and mark post/removal.
2. Powerful equipment for selecting substances and passing on their property.
3. Entire maintenance for buses including element pins, inter-sheet terminal, module ports, and wires.
4. Schedule of equipment and Electrical Rules ensure report.
5. Netlist output is to furnish all fashionable PCB layout tools.

## 5 Results and Discussion

### *5.1 Schematic Diagram*

The power of its construction has allowed us to assimilate primary conventional chart based imitation and now with Proteus VSM—interactive circuit replication keen on the plan surroundings. Figure 4 shows the schematic diagram of shortest path identification using greedy in Proteus simulation.

### *5.2 Simulation Output*

Figure 5 the simulation yield of shortest path identification using a greedy algorithm for Proteus simulation. It provides the whole administer of drawing appearance in turns of row thickness, pack style, and standard in addition to fonts.

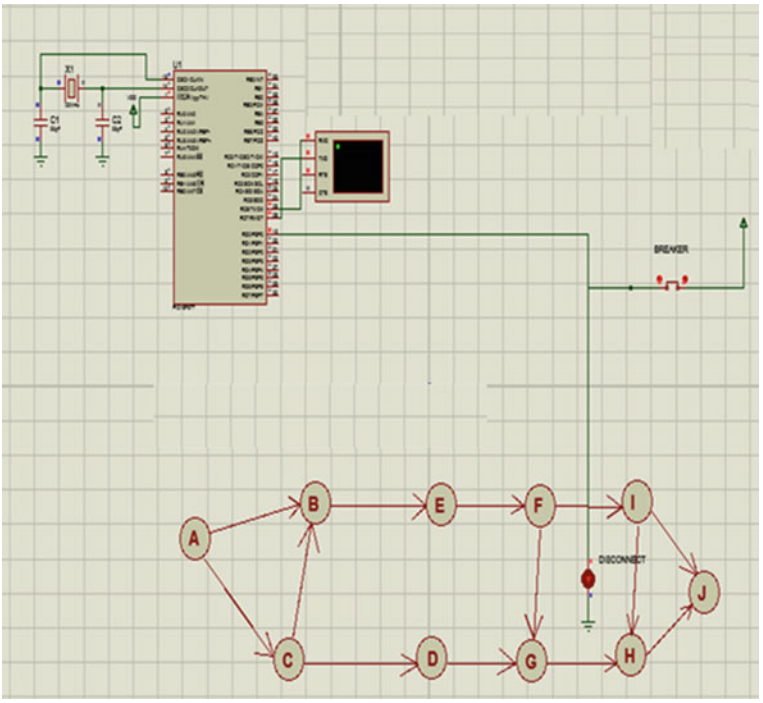



**Fig. 4** Schematic figure of shortest path using greedy

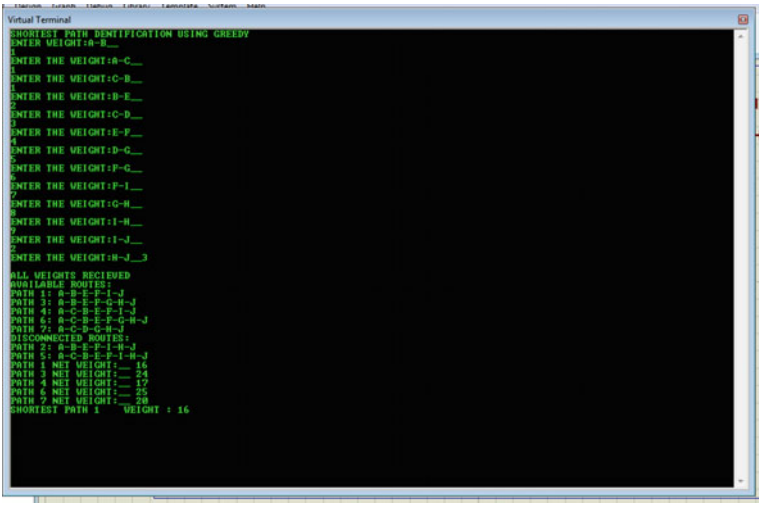



**Fig. 5** Simulation output of shortest path classification using greedy algorithm

**Table 2** Power consumption

| Path | Weight | Power loss transmit mW | Power loss receive mW |
|---|---|---|---|
| Path1 | 37 | 55.5 | 74 |
| Path2 | 44 | 66 | 88 |
| Path3 | 42 | 63 | 84 |
| Path4 | 44 | 66 | 88 |
| Path5 | 51 | 76.5 | 102 |
| Path6 | 49 | 73.5 | 98 |
| Path7 | 30 | 45 | 60 |

## *5.3 Power Analysis of Greedy Algorithm*

The amount of power utilization for communicating through dissimilar nodes inside the range is listed in Table 2.

## *5.4 Simulation Waveform of Greedy Algorithm*

The relationships connecting the numbers of nodes and initial energy capacity using a greedy algorithm, and the results of the greedy algorithm is shown in Fig. 6 In this figure, it is obvious that the network life span increases along with the increasing initial energy capacity.

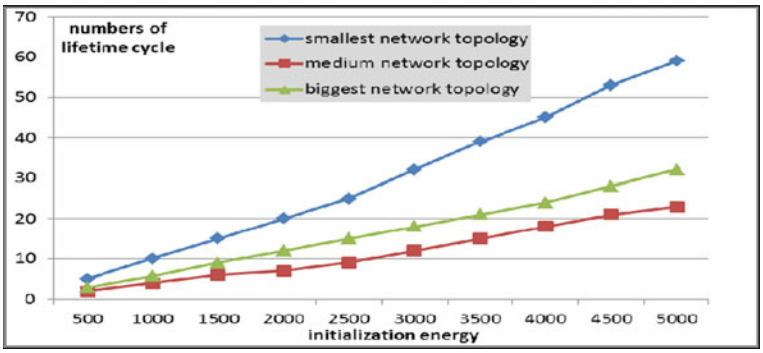



**Fig. 6** Numbers of nodes and initial energy capacity using a greedy algorithm

**Table 3** Power analysis of ant colony algorithm

| Path | Net weight | P Power loss/transmit mW | Power loss/receive mW |
|---|---|---|---|
| Path 1 | 187 | 76.4 | 110 |
| Path 2 | 129 | 63.5 | 87.8 |
| Path 3 | 56 | 45.3 | 63 |
| Path 4 | 131 | 64.7 | 83.5 |
| Path 5 | 131 | 64.8 | 83.5 |
| Path 6 | 86 | 55.9 | 75.2 |

## *5.5 Power Analysis of ACO*

The sum of power utilization for communicating with dissimilar nodes surrounded by the variety is listed in Table 3.

## *5.6 Simulation Waveform of Greedy Algorithm and ACO*

Figure 7 shows the relations connecting the system lifetime initial power capacities along with different algorithms; it is clear that the net lifetime increases and extends the life series of the wireless sensor network since it manages the energy and control management.

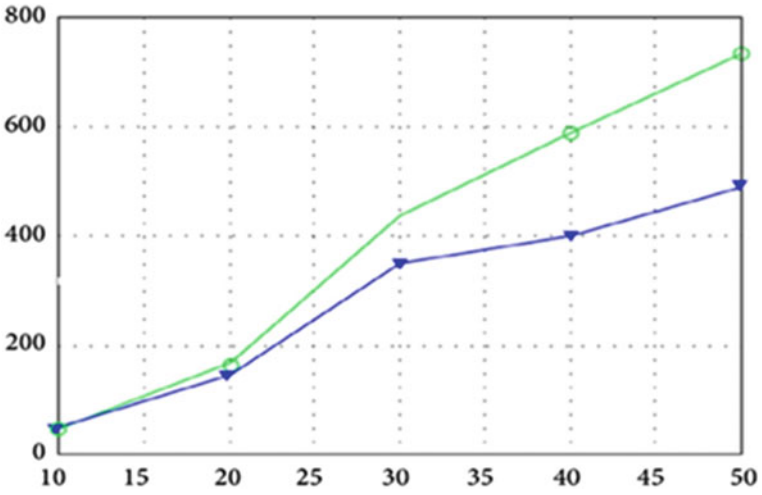



**Fig. 7** Relationships between the network lifetime and Initial energy capacity along with different algorithms. Ant colony algorithm, Greedy algorithm

# 6 Conclusion and for Future Work

## *6.1 Conclusion*

This article presents a greedy algorithm that aims at obtaining the best solution that satisfies the given position of constraints and also maximizes a given objective function. To create the difficulty of selecting communication paths as a Traveling salesman crisis and crack it by a greedy algorithm. The manual calculations to find the number of connected covers in shortest path identification using a greedy algorithm. Simulation results can exhibit the shortest path concerning network lifetime and communication reliability.

In this project also, using the competent ant colony algorithm has been functional to the traveling salesman problem to locate the optimal solution in a short time. However, the performance of the ACO algorithm is designed for both high energy efficiency and good power balancing, and maximal energy utilization throughout the network. The relationship between the statistics of nodes and original power capacity using ACO algorithm is evident that the network lifetime increases and extends the life cycle of the wireless sensor network since it manages the energy and power management.

## *6.2 Scope for Future Work*

In the future, another proposed ACO algorithm will be regularly considered to recover the network lifetime and communication reliability, for example, the active adaptation of multiple routing problems that are essentially multiple paths multiple destinations of the original TSP; this is similar to vehicle routing problems. For these problems, numerous routes are considered, which makes them closer to real-world applications.

# References

1. Kakhandki AL, Hublikar S, Priyatamkumar (2018) Energy efficient discriminatory hop selection optimization toward maximize lifetime of wireless sensor networks. Alexandria Eng J 57(2):711–718
2. Akbas A, Yildiz HU, Tavli B, Uludag S (2016) Joint optimization of transmission power level and packet size for WSN lifetime maximization. IEEE Sens J 16(12):5084–5094
3. Deng W, Xu J, Zhao H (2019) An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem. IEEE Access 5(7):20281–20292
4. Lin Y, Zhang J (2012) An ant colony optimization approach used for maximizing the lifetime of heterogeneous wireless sensor networks. IEEE Trans Syst 42(3):408–420
5. Sun Y, Dong W, Chen Y (2016) An improved routing algorithm based on ant colony optimization inside wireless sensor networks. IEEE Commun Lett

6. Bagula A, Mazandu K (2008) Energy constrained multipath routing in wireless sensor networks. In: Proceeding of the 5th international conference on ubiquitous intelligence and computing, Oslo, Norway, pp 453–467
7. Chang CT, Chang CY, Zhao S, Chen JC, Wang TL (2016) SRA: a sensing radius adaptation mechanism for maximizing network lifetime in WSNs. IEEE Trans Veh Technol 65(12):9817–9833
8. Fonseca R, Gnawali O, Levis K (2007) Four-bit wireless link estimation. In: Proceedings of the 6th workshop taking place hot topics in networks (Hot Nets VI), Atlanta, GA, USA
9. Yetgin H, Cheung KTK, El-Hajjar M, Hanzo L (2014) Cross-layer network lifetime maximization in interference-limited WSNs. IEEE Trans Veh Technol 64(8):3795–3803
10. Wang H, Agoulmine N, Ma M, Jin Y (2010) Network lifetime optimization in wireless sensor networks. IEEE J Sel Areas Commun 28(7):1127–1137
11. Cohen K, Leshem A (2010) A time-varying opportunistic approach to lifetime maximization of wireless sensor networks. IEEE Trans Sig Process 58(10):5307–5319
12. Lin K-Y, Wang P-C (2010) A greedy algorithm in WSNs intended for maximum network lifetime and communication reliability. In: IEEE 12th international conference on networking, sensing and control. Howard Civil Service International House, Taipei, Taiwan
13. Younis M, Senturk I (2012) Topology management techniques for tolerating node failures in wireless sensor network. A review. Comput Netw pp 254–283
14. Imon SK, Khan A, Di Francesco M, Das SK (2014) Energy-efficient randomized switching for maximizing lifetime in tree-based wireless sensor networks. IEEE/ACM Trans Netw 23(5):1401–1415

# Deep Ensemble Approach for Question Answer System


**K. P. Moholkar and S. H. Patil**

**Abstract** Researches on question answering systems has been attracting significant research attention in recent years with the explosive data growth and breakthroughs in machine learning paradigm. Answer selection in question answering segment is always considered as a challenging task in natural language processing domain. The major difficulty detected here is that it not only needs the consideration of semantic matching between question answer pairs but also requires a serious modeling of contextual factors. The system aims to use deep learning technique to generate the expected answer. Sequential ensemble approach is deployed in the proposed model, where it categorically boosts the prediction of LSTM and memory network to increase the system accuracy. The proposed model shows a 50% increase in accuracy when compared to individual systems with a few number of epochs. The proposed system reduces the training time and boosts the system-level accuracy.

**Keywords** Question answer system · Deep neural network · LSTM · Memory · Network ensemble · CatBoost

## 1 Introduction

In general, people tend to seek information through conversation, Internet, books, etc. Due to information overload, the process of extracting the required answers using machine learning techniques has become a challenging task. The task become more complicated when the users have started to post different types of questions like Wh, short question, factoid, reasoning questions, counting, etc. A system is always expected to learn facts and apply reasoning to discover new facts that can assist in answering the given question. Extraction and identification of suitable answer

K. P. Moholkar (✉)
Bharti Vidypeeth (Deemed to be University), Pune, India

S. H. Patil
College of Engineering, Bharti Vidypeeth (Deemed to be University), Pune, India
e-mail: shpatil@bvucoep.edu.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_2

depends on understanding the main intent of a question. This makes the task of building an appropriate model as a challenging task. The conventional method of question system will have data retrieval and handcrafted rules. The performance of traditional systems is restricted as they heavily depend on manually extracted rule. The task becomes difficult because of imbalanced dataset. The QA process can be broken into two parts: information retrieval and reading comprehension. The process of finding the document containing the answer for the raised question is known as information retrieval. Reading comprehension deals with identifying answer from a given the passage. However, the recent schemes on deep learning have shown their potential in improving the performance of QA system. Deep learning models require a substantial amount of training. GRU, LSTM, and bidirectional LSTM help to handle longer text sequences unrolled during time. Enhancements like attention mechanism and memory networks help the model to identify important features. Dialog systems [8] like chat bots [9], Alexa, and IBM Watson [10] which simulate human conversation are derived from question answer systems. With the increase in digital awareness and reach in rural area, the need of multilingual question answer system has started arising. Availability of multilingual QA corpus is a challenging task. Until recently, multilingual QA system was dependent on machine translation. Identifying appropriate answers and framing syntactically and semantically correct sentence require consideration of language-specific grammar, which is also difficult task. This would further affect the accuracy of system. Therefore, advancements in QA system are the need of the hour. Ensemble models help to combine the strengths of individual models to enhance the accuracy of system. The major contribution of the paper is an ensemble model which categorically boosts the combined predictions of LSTM model and memory model. The ensemble model is designed to boost the accuracy of question answering system capable of handling different question categories. The proposed research work has been arranged as follows: Section 2 provides a brief review of literature. Section 3 explains the overview of the proposed question answering system. Section 4 presents experimental setup. Section 5 illustrates the results and discussion, and Sect. 6 presents the conclusion.

## 2 Literature Review

Abishek et al. [1] proposed a CNN—LSTM model for identifying duplicate questions using Quora dataset. The author proved that identifying duplicate questions before the classification enhances the performance of model tremendously. Bae et al. [2] employed word weighting method for question classification. The author proposed mechanism to identifying positive and negative words, and positive words having negative sentiments, and vice versa for query processing. Razzaghnoori et al. [3] experimented with recurrent neural network (RNN) for extracting feature vector mechanism and employed classical support vector machine (SVM) for query classification. Hong et al. [4] have proposed a model that constructs a binary tree structure to take care of composite reasoning implied in the language. The recursive

grounding tree (RVG-TREE) parses the language and then performs a visual reasoning along the tree in a bottom-up fashion. Liu [5] proposed a Siamese architecture that compromises of multilayer long short-term memory (LSTM) network and convolutional neural networks (CNN). The author uses concept interaction graph for matching long text. The graph vertex acts as a concepts, and interaction level act as edges. The model aggregates the matching signals through a graph convolutional network which checks similarity relationship between two articles. Ullah et al. [6] proposed an attention model that emphasizes on the importance of context and time that helps to provide contextual information regarding the question while generating the answer representations. In 2017, Yue et al. [7] developed a dynamic memory networks to carry out "textual QA." The model takes the inputs that are processed to take out hierarchical and global salient features simultaneously. Consequently, they were deployed to formulate numerous feature sets at every interpretation phases. Experimentation was performed on a public textual Q&A dataset using without and with supervision approach from labels of constructive details. Finally, when distinguished with preceding works, the developed technique demonstrates improved stability and accuracy.

## 3 Proposed System

### *3.1 Long Short Term Memory*

LSTM helps to defend the error that is backpropogated through time. LSTM is the memory element in the model. A LSTM cell consists of neural network with three different activation functions. LSTM cell has three basic gates which are neural networks with sigmoid activation function known as input gate '*i*', forget gate '*f*', and output gate '*o*'. Neural net with Tanh activation is known as candidate layer '*c*'. Output is a current hidden state id denoted by vector '*h*', and memory state denoted by vector '*c*'. Forget layer '*F*' decides which information to discard. Inputs to the LSTM cell at each step are $X$ the current input and $H$ the previous hidden state, and C is the previous memory state. Outputs from the LSTM cell are $H$ the current hidden state and $C$ the current memory state. The following equations demonstrate the working of LSTM layers.

$$f_t = \sigma(W_f.[h(t-1), x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_f.[h(t-1), x_t] + b_i) \tag{2}$$

$$\tilde{C} = \tanh(W_C.[h(t-1), x_t] + b_C) \tag{3}$$

$$C_t = f_t * C(t-1) + i_t * \tilde{C}) \tag{4}$$

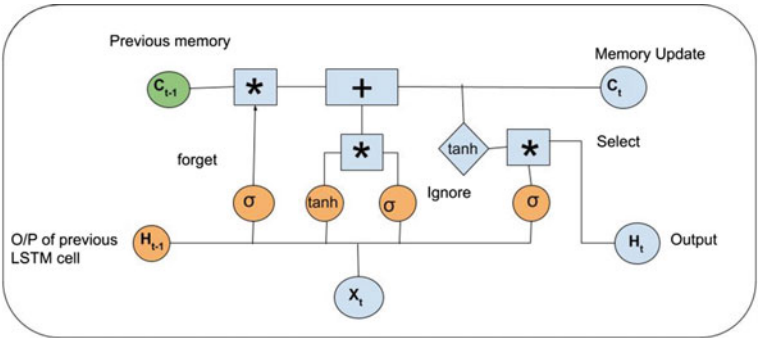



**Fig. 1** LSTM cell

$$o_t = \sigma(W_o.[h(t-1), x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

Figure 1 demonstrates the working of LSTM cell. The cell compares its memory with previous output of LSTM cell to identify the importance and reads a current word considering previous.

## *3.2 Encoder–Decoder Model*

For a QA system to be efficient, inference mechanism and memory elements are required to memorize long sequences with context. These long sequences serve as knowledge for effective prediction of answers. This model decodes the context and question in fixed length vector and produces the expected answer. This model is also known as encoder–decoder model. Consider a network with $M = \mu_1, \mu_2, \ldots, \mu_n$ memory elements, $I$ input feature of vector of size $e$, $u$ update parameter for updating old memory with present input, $O$ the output feature map, and $R$ the expected response. The encoder processes the input feature and produces a two-dimensional output matrix. The model takes input as sentence and stores in available memory $\mu$. The input length depends on the number of memory cells in the layer. A new memory is stored in $u$ module without changing the old memory $\mu$. LSTM layer acting as a decoder takes input vector, the time steps, and features to produce a decoded sequence, i.e., answer (Fig. 2).

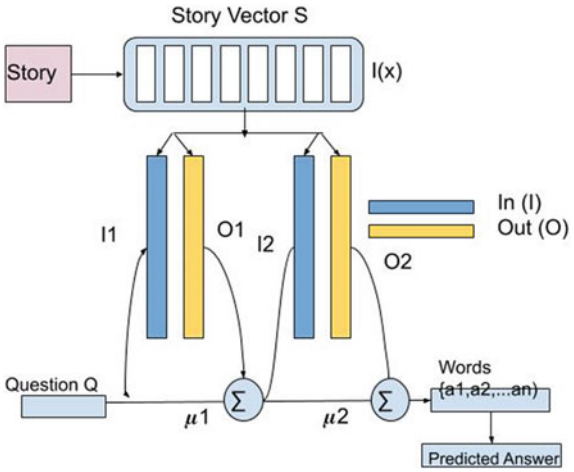



**Fig. 2** Memory network

**Algorithm 1:** Memory Cell

**Data:** input text character/word/sentence
**Result:** Memory Network Algorithm
initialization;
**while** *not at end of sentence* **do**
  Encode x → I(x);
  **while** *not at end of X* **do**
    Update Memory
    $\mu_i = u(\mu_i, I(x), \mu), \exists i$
    Compute $O = O(I(x), \mu)$
    end of X
    Decode $o : r = R(o)n$;
  **end**
**end**

## *3.3 CatBoost*

CatBoost is an implementation of gradient boosting, which uses binary decision trees as base predictors [12]. Existing gradient boosting algorithms like random forest, Adaboost, and Xgboost suffer from prediction shift problem. Boosting algorithm relies on target of all training examples which causes prediction shift. Traditional approach [14] converts categorical features to statistical model leading to target leakage. CatBoost algorithm provides solution to both the said issues by employing principle of ordering. CatBoost is based on gradient boosted decision trees. Decision

trees are built repetitively during training phase. Each successive tree is built with reduced loss compared to the previous trees. When the algorithm identifies over fitting, it stops construction of new tress.

**Algorithm 2:** Categorical Boosting

```
Data: input text character/word/sentence
Result: tree
Preliminary calculation of splits.;
Choosing the tree structure.;
while not over-fitting do
    Divide objects into disjoint ranges;
    Split data into bucket;
    Select tree structure;
    Calculating values in leaves.;
    while not over-fitting do
        Calculate penalty function;
        min(penalty)
        Select split
    end
    Calculate Score
end
```

The CatBoost algorithm is good in handling categorical values and performs best when used with default parameters, and parameters can be tuned for optimized results. Though the algorithm is robust and easy to use, it requires considerable time to train and optimize the results. CatBoost takes more time to train a system with numerical features. It does not support sparse matrices.

## *3.4 Ensemble Model*

A deep neural network (DNN) learns complex nonlinear relationship in data. Bias, variance, and noise are prediction errors. While noise cannot be handled by algorithmic approach, bias and variance can be reduced by proper choice of model. Every time a DNN is trained, different version of mapping function is learnt. This stochastic nature affects the performance of model. The DNN suffers from low bias and high variance. The proposed model uses ensemble approach to reduce the problem of high bias. Sequential ensemble approach is used in the proposed system to reduce bias error. This approach combines predictions from base learners and applies categorical boosting to enhance the accuracy of system. Several instances of the same base model are trained sequentially to train the LSTM learner with previous weak learners. Figure 3 shows the working of the model. Long short-term recurrent neural network and memory networks are used as base learners. Categorical boosting is done by CatBoost algorithm which is used as meta learners to enhance the accuracy of ensemble model. The proposed model works in two phases. The learning

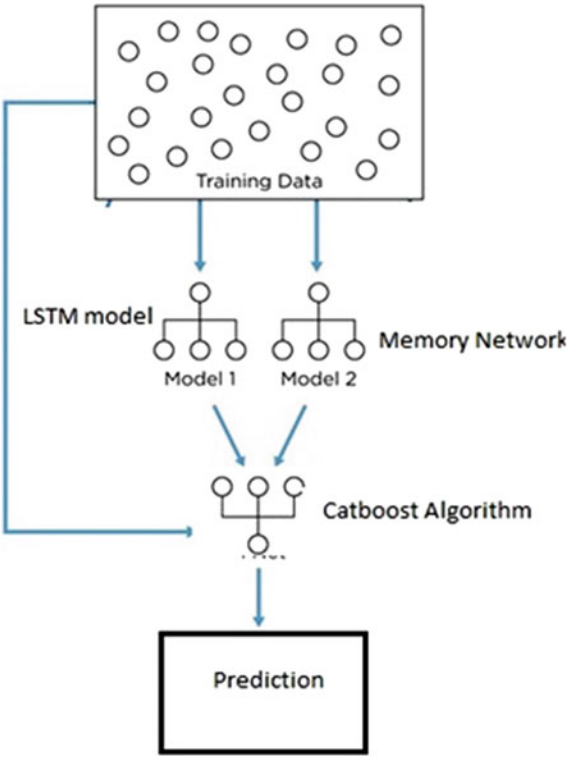



**Fig. 3** Proposed model

algorithms for LSTM are based on minimizing the error rate. The first phase uses data subsets to produce results. The results generated by first phase are combined together, and performance boost is achieved by categorical boosting (Fig. 3).

## 4 Experimental Setup

LSTM and memory networks are trained in full supervision by providing context, question, and expected answers to the system. Each model is tested in isolation for accuracy and loss parameters. Evaluation of the proposed system is done on babi question answering dataset. The dataset consists of 20 different question answering tasks. The models are trained with 40 epochs and batch size of 32. The model is trained on 1000 samples. A dropout of 0.3 is applied to avoid over fitting. The efficiency of a QA system is judged by the correctness of answer predicted. Precision and recall are traditional factors for judging a classification problem. Apart from these, mean reciprocal rank (MRR) is the ability of a QA system to select appropriate answer for

a given question from the ranked list of answers. If suitable answer is not found, then rank becomes zero. The proposed model was evaluated on precision, recall, and f1 measure.

$$\text{Precision (Micro)} = \frac{\text{No. of correctanswers for Question } q}{\text{answers retrieved for Question } q} \tag{7}$$

$$\text{Recall (Micro)} = \frac{\text{Correct answers retrieved for Question } q}{\text{gold standard answers retrieved for } q} \tag{8}$$

$$\text{Precision} = \text{Average of Precision for all questions} \tag{9}$$

$$\text{Recall} = \text{Average of Recall for all questions} \tag{10}$$

$$\text{F1 Score (Micro)} = \text{Harmonic mean of Precision (Micro) and Recall (Micro)} \tag{11}$$

$$\text{F1 Score} = \text{Harmonic Mean of Precision and Recall} \tag{12}$$

$$\text{MRR} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{\text{answerrank}_i} \tag{13}$$

## 5 Results and Discussion

From experiments, it was observed that each classifier requires different settings of hyperparameters for producing best model. With fixed hyperparameter setting, individual performance of classifier varies a lot. The proposed ensemble boosting helps to overcome this bottleneck. Considerable amount of time needs to be invested to identify optimal hyperparameters. In cases where single classifier is not sufficient to correctly predict the result, the proposed approach provides a better alternative. Metric classification report indicates that proposed ensemble model performs better than individual models. The accuracy of LSTM model is obtained around 60% and memory network model around 60% for 40 epochs. The accuracy of proposed ensemble model obtained is 95.45%. The f1 score of proposed model is 93%. A considerable boost in accuracy is observed in the result (Table 1).

**Table 1** Comparing different models on precision, recall, and f1 score

| Model | Precision | Recall | F1 score |
|---|---|---|---|
| LSTM | 51 | 50 | 50 |
| Memory network | 37 | 38 | 37 |
| Proposed model | 91 | 95 | 93 |

The ROC curve is plotted with true positive rate on *y*-axis and false positive rate on *x*-axis. The ROC curve demonstrates the probability distribution of system. Classifiers that produce curves closer to the top-left corner indicate a better performance. The ROC curve and FPR FNR curve indicate that the performance of proposed model is stable. FNR indicates the miss rate for given data (Figs. 4 and 5).

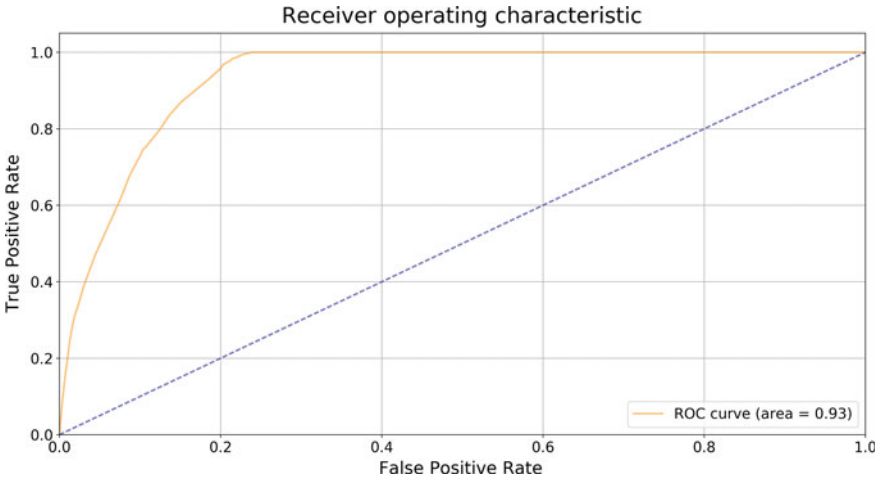



**Fig. 4** ROC curve

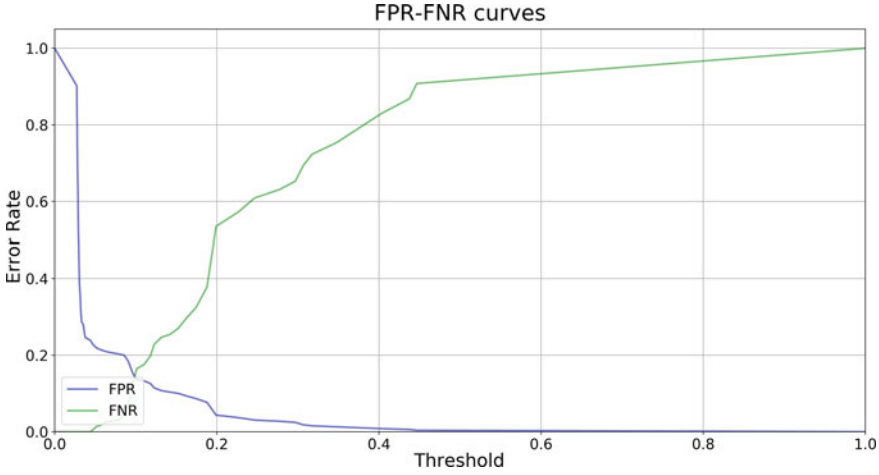



**Fig. 5** FPR/FNR curve

## 6 Conclusion

Experiments explore proposed model for question answering (QA) task. The experiments show that using LSTM and memory network as base learners and boosting the results with CatBoost algorithm shows considerable improvement in prediction accuracy. Evaluation of proposed system is done on babi dataset which contains 1000 question answer tasks. Overcoming the problem of bias was handled successfully in the system. There are many future enhancements for the said system. Choice of base learners and combining traditional machine learning algorithm would make it computationally efficient. Testing system with multilingual data would be a challenging task.

## References

1. Abishek K, Hariharan BK, Valliyammai C (2019) An enhanced deep learning model for duplicate question pairs recognition. In: Soft computing in data analytics. Springer, Singapore, pp 769–777
2. Bae K, Ko Y (2019) Efficient question classification and retrieval using category information and word embedding on cQA services. Springer J Intell Inf Syst 53:27–49
3. Razzaghnoori M, Sajedi H, Jazani IK (2018) Question classification in Persian using word vectors and frequencies. ACM J Cogn Syst Res 47(C):16–27
4. Hong R et al (2019) Learning to compose and reason with language tree structures for visual grounding. IEEE Trans Pattern Anal Mach Intell
5. Liu B, Zhang T, Niu D, Lin J, Lai K, Xu Y (2018) Matching long text documents via graph convolutional networks. arXiv:1802.07459v1 [cs.CL] 21 2018
6. Ullah A, Xiao H, Barker T (2019) A study into the usability and security implications of text and image based challenge questions in the context of online examination. Springer Educ Inf Technol 24(1):13–39
7. Yue C, Cao H, Xiong K, Cui A, Qin H, Li M (2017) Enhanced question understanding with dynamic memory networks for textual question answering. Expert Syst Appl 1(80):39–45
8. Manning C, Text-based question answering systems, p 7. http://web.stanford.edu/class/cs224n/handouts/cs224n-QA-2013.pdf
9. Quarteroni S (2007) A chatbot-based interactive question answering system. In: 11th Workshop on the semantics and pragmatics of dialogue, p 8390
10. Murdock JW (Guest Editor) (2012) This is Watson. IBM J Res Dev 56(3/4)
11. Qu C, Ji F, Qiu M, Yang L (2018) Learning to selectively transfer: reinforced transfer learning for deep text matching. arXiv:1812.11561v1 [cs.IR] 30 Dec 2018
12. Prokhorenkova L et al (2018) CatBoost: unbiased boosting with categorical features. In: Advances in neural information processing systems
13. Dorogush AV, Ershov V, Gulin A (2018) CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363
14. Micci-Barreca D (2001) A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. ACM SIGKDD Explor Newsl 3(1):27–32
15. Cestnik B et al (1990) Estimating probabilities: a crucial task in machine learning. ECAI 90:147–149
16. Moholkar K, Patil S (2019) Hybrid CNN-LSTM model for answer identification. Int J Recent Technol Eng (IJRTE) 8(3). ISSN: 2277-3878

# Information Sharing Over Social Media Analysis Using Centrality Measure


**K. P. Ashvitha, B. Akshaya, S. Thilagavathi, and M. Rajendiran**

**Abstract** Instagram and Twitter are popular social media in India. This exploration plans to comprehend the client's behavior and strong/weak relationship inside the social media and to find the range of interaction in the social network. UCINET is an analytic tool for social media networks which examines the client highlights and floating inspiration of these four environments. To measure the centrality three markers such as degree centrality, closeness centrality and betweenness centrality are widely used. Clients obtained the most astounding estimation of centralities are mostly used in social media by people, and the least centrality is used the least. The recurrence to utilize four web-based lives is comparable as per the results of three centralities.

**Keywords** Social media · Centrality · Social network analysis

## 1 Introduction

Immense SNA originates from social science, network analysis, and graph theory. Network analysis produces a result for the problems which are in the form of a network generally in the graph structure. There are many analytical tools to analyze this network and visualize it in the form of a graph. The main purpose of designing social network analysis is to work on groups especially on the social network rather than on individual data. SNA is used in the field of business for improving communication flow, law enforcement agencies to find forgeries network, social networks such as Facebook to find a friend of friends, and network operators such as telephone, cable, and even in mobile networks. The relationship between two nodes can be unidirectional or bidirectional. For example, Facebook is unidirectional where only one person is enough to accept the request but in the private account of Instagram, both of them need to send a follow request and accept it mutually, so Instagram is bidirectional.

K. P. Ashvitha (✉) · B. Akshaya · S. Thilagavathi · M. Rajendiran
Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_3

A social group can be of two types such as realist and nominalist. The social network comes under the group of nominalists and not under realist because the actors have a relationship among themselves. There are two methods for analyzing social network such as complete network and ego network-based analysis. Traditional survey is used to analyze the ego network and also involves evaluating the parameters such as average, size, min, and max of a person network. Whereas complete network analysis is used to produce the relationship between the actors such as trustworthy friends in a connected network, and it is also used to measure centrality.

UCINET is a user-friendly software system for analyzing the social network. It also provides several analytical technologies such as clustering, centrality, and statistical analysis. UCINET accepts a large number of inputs in the form of an excel file using a data editor known as DL editor. Excel can either be created in the DL editor or imported from the system. They are different features in which data can be arranged are Full matrix, Nodelist, Edgelist, Edgelist23, etc., Here, the Full matrix feature is used in which data are arranged as actor-by-actor adjacency matrix in which $X_{ij}$ shows the strength of relation from actor $i$ to actor j. Full matrix is mainly used for small and dense data sets.

NETDRAW enables the client to delineate and connect utilizing drawing highlights such as color, size, and shapes. Tie strength is used to differentiate the different relationships in the relation using the attribute. Thus, the result of the NETDRAW can be stored in different formats such as Windows metafile and jpeg format. It is also used for clustering and measuring centrality. UCINET produces two types of output. At first, the output log opens the text file in the notepad gets displayed default. Once the UCINET is closed, the output log file gets deleted so it is necessary to store the output log. The other way of viewing data is through UCINET which creates UCINET data file which contains results. This file will not get deleted when the UCINET is closed, so the user can view the output whenever required.

## 2 Related Works

In leadership development, one of the challenging tasks is evaluation of social networks. Mathematics and perception-based approach such as social network analysis (SNA) is utilized to frame the links for individuals, objectives, benefits, and different objects inside a large system [1]. SNA is utilized to expand the attention toward intensity of systems to leaders, to additional connections, associations, and finally reinforce the system limit altogether. The social organization conducted an investigation to evaluate all of its initiative improvement efforts, including the Emerging Leaders programme. The structure for administration systems supports the user by comprehending the work and help to choose and decide how and when to utilize SNA for assessment and capacity-building tool. An unpredictable system module is worked to speak the mind complex item structure, utilizing short-term esteemed intuitionist logic figure out the edge weights of the system demonstration and provides relationship qualities between measures [2]. Due to this, UCINET

is utilized for item module segment. The module is based on mutual relationship of complex products. The instrument of informal community examination utilized UCINET to exchange the co-word network into a photo of the keyword system and make an examination of it. They break down the arrangement as well as connections characteristics formed through various social entities [3].

Once the social network research organization examines to perceive the generality of its structural approach, helpful applications in an extensive variety of experimental circumstances moved toward becoming possible. SNA study is one of the few social endeavors in which individuals impact each other such that they all cooperate to assemble an aggregate assemblage of knowledge [4]. In the web condition, the hyperlink quantity checks and quantity of webpages are utilized to measure the affect factor. Self-references are supplanted without anyone else joins, i.e., connections inside the sites, and references are supplanted through in-joins, i.e., incoming links from other sites. There are several approaches to discover the effect of dairy, paper, and web locales, and so on. The proposed framework will discover the affect factor of understudies E-Dissertation through h-record and associated connections of E-Dissertation spoken to pictorially by utilizing UCINET programming [5].

The centrality of the venture administration group is very high, which implies it has the greatest power in the correspondence organization, yet that will influence the general correspondence productivity [6]. In migration estimation, social networks have denoted a significant takeoff of understanding. It is not just a component as the movement procedure is spotted. This extraordinary issue arises as migrants are implanted between individual connections and primary subjects, directions of no less than two distinct country states. In this way, their lives span crosswise over borders. The creators appear among others that migrants and non-migrants have very comparative talk networks and participation on different measurements, yet they vary in the ethnic composition of their systems [7].

SNA objective depends on scientific investigation, commonly on estimations that are defined on the 60s and the present created strategies. Estimations and relations between them will be broke down, utilizing the current programming for informal community examination, for example, UCINET and ORA. This related work is finished on numerous means, while for information, investigation is utilized DBLP dataset [8]. Utilizing a self-administrated survey, 210 individuals are requested to give a reply for the study utilizing the accommodation examining strategy [9]. Individuals spend more than 33% of their waking day expending social media. However, with the notoriety of social media sites, a few organizations are utilizing informal communication destinations to help the production of brand networks [10].

The fundamental purpose of this article was the investigation of co-initiation in a particular meeting and the connection of these co-creators with paper procedures. Research through poll study and meetings to gather data and manufacture a system of connections, utilizing a blend of subjective quantitative examination, with the guide of the UCINET social organize examination programming to break down the impact level of relationship writes and comparing system impacts, accomplish the graphical portrayal of the system topological structure [11].

In addition, exploration estimated the grade in light of informal community investigation techniques, dissects the networks into strong subcategories along with group calculation and inner circle. The point is to recognize the impact of online networking promoting on brand loyalty of the shoppers, given that the idea is accepting and expanding consideration from advertising the scholarly community and experts. Data was gathered through the organization of an organized survey with an example of 338 individuals who were online networking users [12]. One question asked whether the respondent was utilizing web-based life more than once in seven days. The respondent was utilizing web-based life more than once in seven days, and they took after no less than one brand via web-based networking media. The questionnaire was created to quantify measure brand loyalty in social media.

## 3 Data Analysis

The analysis is experimented with 25 active participants in social media such as Instagram and Twitter. Some of the members were interviewed and constructed the matrix of interaction. In view of the meeting, information of connection among actors has been collected. The parameters were exposed to the contiguousness network, which is a sort of connection between clients and server who are hubs of social systems. The information was investigated based on the product UCINET 6, which is the contiguousness network input parameter. The outcomes acquired and in addition to their understanding are identified with the accompanying measures such as centrality (degree, betweenness, closeness) grouping coefficient, thickness, achieve, geodesic separation, and eigenvector [13].

Social information is represented with charts called sociograms. A different sociogram is normally worked for each social connection under study. Sociologists as a rule get them through public opinion surveys and meet with the people. Also, the investigation of human social structures is normally done by a socialist external to society. Social network analysis organizes study can be utilized as a part of a notoriety framework that considers the social measurement of notoriety. The combination of complementary strategies that utilization diverse parts of the communication and social relations, enables the specialist to compute notoriety esteems at various phases of its knowledge of the society[14] (Table 1).

This research intends to comprehend the client's behavior (features) and strong/ties inside the social network and if clients have been drifted on two social media. Analytic tool for social media networks UCINET is utilized in the research which examines the client highlights and floating inspiration of these two environments. To measure the centrality, three markers such as degree centrality, closeness centrality, and betweenness centrality are used [15].

A. *Analysis of Centrality.*
A total of 20 samples is used to create two social networks in view of two platforms (Instagram and Twitter). The created systems depended on the discussion

**Table 1** Degree of interaction

| Degree of interaction | Explanation |
|---|---|
| 0 | No interaction |
| 1 | Interact with others every 2 days |
| 2 | Interact with others every 5 days |
| 3 | Interact with others every week |
| 4 | Interact with others every two weeks |
| 5 | Interact with others every month |

data which is equally to the level of communication. Quantities of centralities are contrasted to decide if the actors drifted or not (Figs. 1, 2 and 3).

B. *Degree Centrality.*
To characterize the incident quantity of a node along the lines, degree centrality is utilized. The most potential node among nodes can be defined through degree centrality. In order to obtain this, direct connection is essential between the nodes, and it is mathematically expressed as

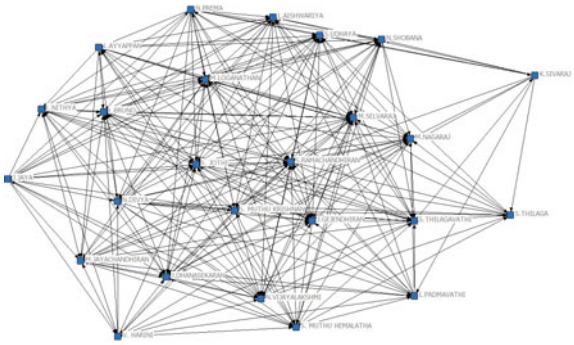


**Fig. 1** Social network generated for 25 members on Instagram

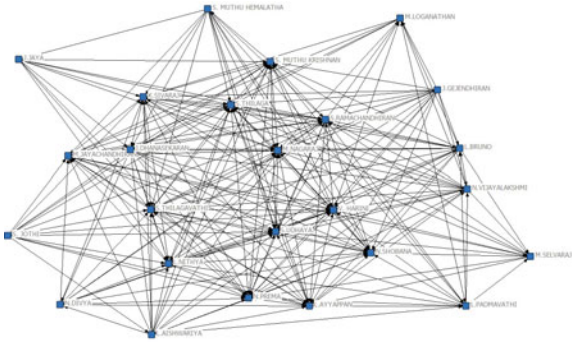


**Fig. 2** Social network generated for 25 members on Twitter

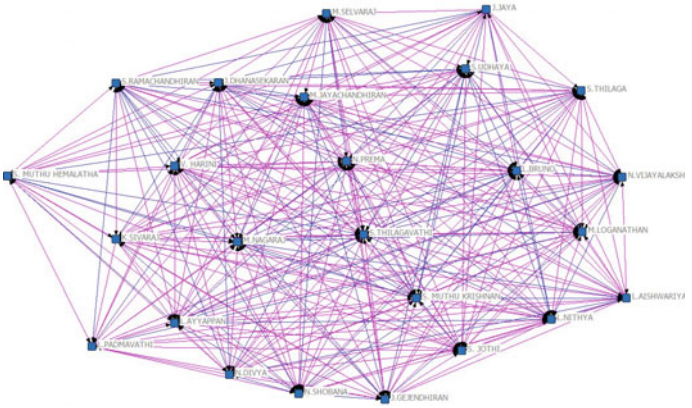



**Fig. 3** Social network generated for 25 members on combined Instagram and Twitter

$$C_D(P_k) = \sum_{i=1}^{n} a(p_i, p_k)$$

The relative measure of degree centrality

$$C'_D(P_k) = \frac{\sum_{i=1}^{n} a(p_i, p_k)}{n-1}$$

The normal degree centrality of Twitter is observed to be 21.000 by L. Ayyappan, and Instagram is observed to be 21.000 by S. Jothi, and a combination of Instagram and Twitter is observed to be 39.000 by J. Dhanasekaran.

C. *Closeness Centrality.*

The distance between one node and other node is used to represent the closeness centrality. Subsequently, it is used to quantify the magnitude request of a node that is close to another node in the network by computing the shortest path from one node to all nodes in the system as a graph. Mathematically, it is expressed as

The relative closeness centrality is

$$C_C(P_k) = \left[\sum_{i=1}^{n} d(p_i, p_k)\right]^{-1}$$

S. Jothi has a higher incentive in Instagram of 88.889, L. Ayyappan has a higher incentive in Twitter of 88.889, and M. Nagaraj has a higher incentive in both Instagram and Twitter of 100.

D. *Betweenness Centrality.*

**Table 2** Comparison value of betweenness and in-betweenness

| Name | Betweenness | Nbetweenness |
|---|---|---|
| S. Ramachandhiran | 46.348 | 8.396 |
| V. Harini | 36.636 | 6.637 |
| N. Prema | 33.08 | 5.993 |
| S. Thilagavathi | 27.008 | 4.893 |
| S. Muthu hemalatha | 24.603 | 4.457 |
| S. Muthu krishnan | 23.778 | 4.308 |
| L. Nithya | 18.122 | 3.283 |
| S. Udhaya | 15.957 | 2.891 |
| L. Ayyappan | 14.762 | 2.674 |
| L. Padmavathi | 13.277 | 2.405 |
| K. Sivaraj | 10.418 | 1.887 |
| N. Shobana | 8.409 | 1.523 |
| S. Thilaga | 8.202 | 1.486 |
| M. Nagaraj | 6.763 | 1.225 |
| J. Dhanasekaran | 5.56 | 1.007 |
| N. Vijayalakshmi | 5.145 | 0.932 |
| M. Jayachandhiran | 4.494 | 0.814 |
| M. Loganathan | 2.555 | 0.463 |
| N. Divya | 1.834 | 0.332 |
| L. Bruno | 1.833 | 0.332 |
| L. Aishwariya | 1.229 | 0.223 |
| S. Jothi | 0.894 | 0.162 |
| M. Selvaraj | 0.726 | 0.132 |
| J. Jaya | 0.365 | 0.066 |
| J. Gejendhiran | 0 | 0 |

Betweenness centrality is used to define the node which lies between other nodes in a network. In the event, if a particular node is the only element which connects two different groups, then that node is considered as vital node in order to keep the cover the network

$$C_B(p_k) = i_{ij}(p_k) = \frac{1}{g_{ij}} * g_{ij}(p_k) = \frac{g_{ij}(p_k)}{g_{ij}}$$

S. Thilagavathi has a value of 27.750 plays an essential in the mediation of the network on Instagram. S. Ramachandhiran has a value of 46.348 plays an essential in the mediation of networks on Twitter. S. Muthu Krishnan has a value of 16.593 plays an essential in the mediation of networks on Instagram and Twitter.

Using Table 2, the graph can be generated using the values shown above.

The above graph gives the generated curve for betweenness and in-betweenness (Fig. 4).

Thus, from the above analysis, calculated value of betweenness and in-betweenness is found (Fig. 5; Table 3).

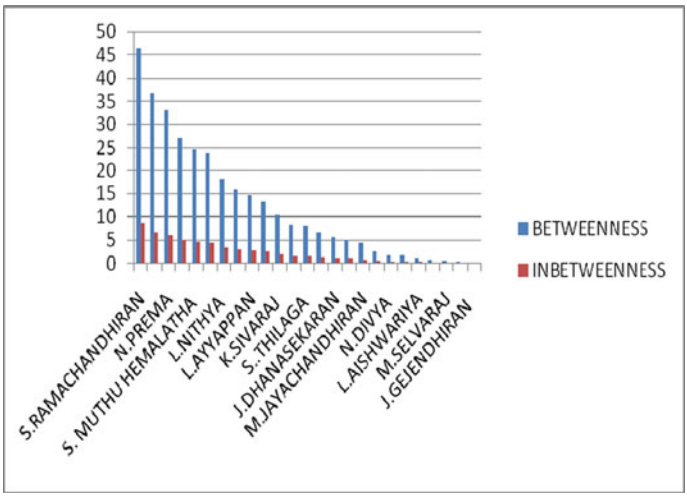



**Fig. 4** Generated graph of betweenness and in-betweenness

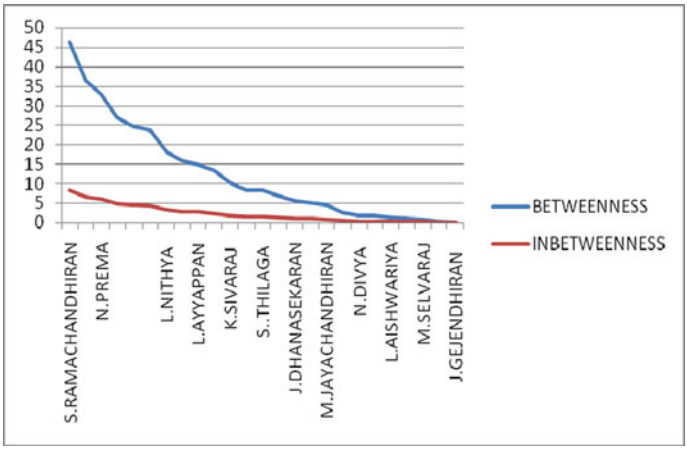



**Fig. 5** Generated curve of betweenness and in-betweenness

**Table 3** Mathematical calculation

| Calculation | Betweenness | In-betweenness |
|---|---|---|
| Mean | 12.48 | 2.261 |
| Std Dev | 12.577 | 2.278 |
| Sum | 312 | 56.522 |
| Variance | 158.171 | 5.191 |
| SSQ | 7848.046 | 257.563 |
| MCSSQ | 3954.286 | 129.775 |
| Euc Norm | 88.589 | 16.049 |
| Minimum | 0 | 0 |
| Maximum | 46.348 | 8.396 |
| N of Obs | 25 | 25 |

## 4 Conclusion

In the proposed work, the results show up that there is a noteworthy impact of implication from Twitter to Instagram. In three centralities, regular clients are practically identical case in two phases. User interface and preference are the two fundamental contrast between the platforms. The outcomes additionally demonstrate the effects of network as various clients select media in view of their companion's utilization. As per research recognition, an inadequacy exists on Twitter. At that point, again the strong bonds existed on Instagram. Numerous members of the research demonstrated contingency behavior since they endeavored to get used to diverse platforms. Therefore, this analytical model is unclear about the drifting behavior. The results of Instagram match the present reality state and associations between performing actors are robust.

## References

1. Hoppe B, Reinelt C (2010) Social network analysis and the evaluation of leadership networks. Leadersh Q 21:600–619. https://doi.org/10.1016/j.leaqua.2010.06.004
2. Zhang N, Yang Y, Zheng Y (2016) A module partition method base on complex network theory. In: IEEE international conference on industrial engineering and engineering management (IEEM), pp 424–428. https://doi.org/10.1109/IEEM.2016.7797910
3. Yanfang L, Nan D (2014) A study of Chinese culture consumption based on co-words analysis and social network. In: IEEE workshop on advanced research and technology in industry applications (WARTIA), pp 551–554. https://doi.org/10.1109/WARTIA.2014.6976319
4. Freeman LC (2004) The development of social network analysis. Book. https://www.researchgate.net/publication/239228599
5. Saraswathi D, Vijaya Kathiravan A, Kavitha R (2013) A prominent approach to determine the excellence of students E-dissertation using H-index and UCINET, pp 1–5. https://doi.org/10.1109/MECO.2013.6601370

6. He QH, Luo L, Li YK, Zhang SQ, Lu YB (2013) Organizational communication of Shanghai expo village project based on social network analysis. In: International conference on management science & engineering, pp 10–18. https://doi.org/10.1109/ICMSE.2012.6414154
7. Bilecena B, Gamper M, Lubbers MJ (2018) The missing link: social network analysis in migration and transnationalism. Soc Netw 53:1–3. https://doi.org/10.1016/j.socnet.2017.07.001
8. Raya V, Çiço (2013) Social network analysis, methods and measurements calculations. In: Mediterranean conference on embedded computing (MECO). https://doi.org/10.1109/MECO.2013.6601370
9. Abzari M, Ghassemi RA, Vosta LN (2014) Analysing the effect of social media on brand attitude and purchase intention: The case of Iran Khodro Company. Soc Behav Sci 143:822–826
10. Laroche M, Habibi MR, Ricard MO, Sankaranarayanan R (2012) The effect of brand image and brand loyalty on brand equity. Comput Hum Behav 28(5):1755–1767. https://doi.org/10.1016/j.chb.2012.04.016
11. Lin W, Dongying L, Haizhang S, Mengying L (2018) A comparative study on the interpersonal network of learning promotion and employment competitiveness. In: International conference on electronics instrumentation & information systems (EIIS). https://doi.org/10.1109/EIIS.2017.8298557
12. Erdogmus IE, Cicek M (2012) The impact of social media marketing on brand loyalty. In: International strategic management conference, pp 1353–1360. https://doi.org/10.1016/j.sbspro.2012.09.1119
13. Boban I, Mujkic A, Dugandzic I, Bijedic N, Hamulic I (2014) Analysis of a social network. In: International symposium on applied machine intelligence and informatics (SAMI), pp 129–132. https://doi.org/10.1109/SAMI.2014.6822391
14. Sabater J, Sierra C (2002) Reputation and social network analysis in multi-agent systems. In: Proceedings of the first international joint conference on autonomous agents and multiagent systems, pp 475–482. https://doi.org/10.1145/544741.544854
15. Chang WL, Li CB, Ting HC (2014) Exploring the drifting behavior on different social media. In: IIAI 3rd international conference on advanced applied informatics, pp 535–536. https://doi.org/10.1109/IIAI-AAI.2014.114

# AndroHealthCheck: A Malware Detection System for Android Using Machine Learning


**Prerna Agrawal and Bhushan Trivedi**

**Abstract** With the boom of malware, the area of malware detection and the use of gadget assist to gain knowledge in research drastically with the aid of researchers. The conventional methods of malware detection are incompetent to detect new and generic malware. In this article, a generic malware detection process is proposed using machine learning named AndroHealthCheck. The malware detection process is divided into four phases, namely android file collection, decompilation, feature mining and machine learning. The overall contributions made in AndroHealthCheck are as follows: (1) designing and implementing a crawler for automating the process of benign files download, (2) collection of unstructured data from the downloaded APK files through the decompilation process, (3) defining a proper mechanism for the feature selection process by performing a static analysis process, (4) designing and implementing a feature mining script for extracting the features from unstructured data collection from APK files, (5) generating a rich homemade data set for machine learning with a huge variety and different flavours of malware files from different families and (6) evaluating the performance of the generated data set by using different types of supervised machine learning classifiers. In this article, the overall architecture and deployment flow of AndroHealthCheck are also discussed.

**Keywords** Malware detection · APK files · Static analysis · Unstructured data · Feature mining · Machine learning

## 1 Introduction

The malware detection domain using machine learning is an emerging area that is being researched extensively these days. The conventional methods used for the

P. Agrawal (✉) · B. Trivedi
Faculty of Computer Technology, GLS University, Ahmedabad, Gujarat, India
e-mail: prerna.agrawal@glsuniversity.ac.in

B. Trivedi
e-mail: bhushan.trivedi@glsuniversity.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_4

detection of malware are more resource and time consuming and are incompetent to detect generic and new malware [1]. The conventional methods used for malware detection include signature-based, resource-based, components-based and permission-based analysis [2], which are not enough to detect the new and generic malware. Machine learning methods acquire on their own from the knowledge given to them as training data and use performed classification on testing data and are highly used for the investigation of the malware [1].

Here, the Android files are used as a proof of concept for the proposed malware detection process. For the proper investigation of the malware, the independent flavours of features and a variety of malware files from different malware families are needed. The existing malware data sets are available and that can be used in machine learning directly, but the Drebin data set is found to be with lesser features and with malware files having less variety of malware families. To generate our data set with independent flavours of features, they were decided to have a huge variety of malware files for better performance in malware detection. It can be directly used by researchers in machine learning.

The overall contributions performed in AndroHealthCheck are as follows: (1) designing and implementing a crawler for automating the process of benign files download [3], (2) collection of unstructured data from the downloaded APK files through the decompilation process [4], (3) defining a proper mechanism for the feature selection process by performing a static analysis process [5], (4) designing and implementing a feature mining script for extracting the features from unstructured data collection from APK files [5], (5) generating a rich data set for machine learning with a huge variety and different flavours of malware files from different families [5] and (6) evaluating the performance of the generated data set by using different types of supervised machine learning classifiers [6]. Using the machine learning classifiers, the performance of the CatBoost classifier is highest with 93.15% accuracy and ROC value of 0.91 [6]. The layout of this article is divided into the following sections. Section 2 describes the overall architectural flow of the AndroHealthChecka—malware detection system. Section 3 describes the overall deployment flow of the AndroHealthCheck—a malware detection system. Section 4 describes the conclusion of the research work.

## 2 Architecture of AndroHealthCheck

This section represents the overall architecture of AndroHealthCheck—the malware detection system. Figure 1 represents the overall architecture of AndroHealthCheck. The AndroHealthCheck is divided into four phases: They are android file collection, decompilation, feature mining and machine learning. In the android file collection [3] phase, the malware and the benign file collection were concentrated as it is the first step for the data collection. In the android file collection phase, the user enters the website URL for downloading the files. This module connects to the website, and after the successful establishment of connection, the website sends the file request

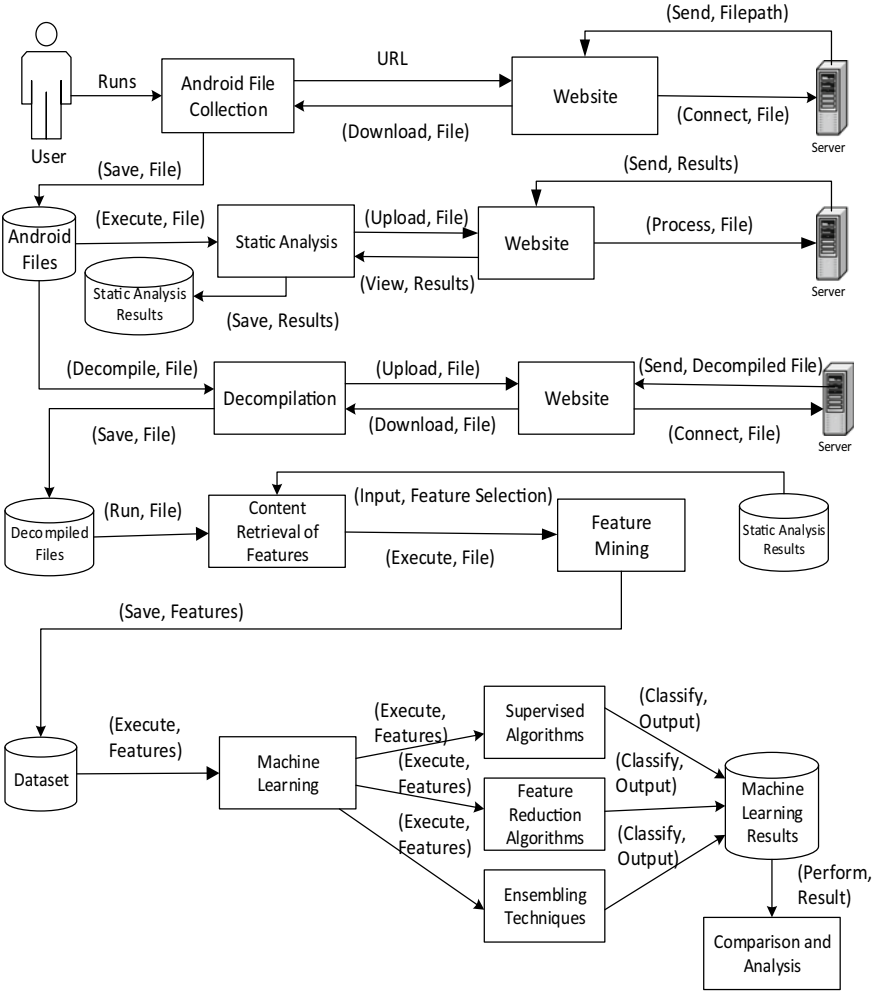



**Fig. 1** Overall architectural flow of AndroHealthCheck

to the server for download. The server responds with the file path to the website, and the file gets downloaded and stored in the physical location. Using the android file collection phase [3], a total of 15,506 files of malware files were downloaded from the world's famous android malware projects like Drebin, Androzoo, AndroPRAGuard, Kharon and Kudoos [3]. To automate the process of benign file downloads, a crawler is developed [3] and downloaded for 4000 benign files [3]. The android files contain an unstructured data format like text files, Java files and.xml files. For extracting the features from the APK files, reverse engineering of these files is necessary. So the decompilation phase [4] collects the unstructured data from the APK files. In the decompilation phase, an APK file is given as input. The APK file is uploaded on the

website and sent to the server for decompilation. The server processes the APK file, decompiles it and sends the decompiled file back to the website in the form of a zip file. The decompiled zip file is saved to a physical location. Using the decompilation phase, the collected malware and benign files are decompiled, and unstructured data like XML and Java files are collected from each decompiled APK file [4].

For mining, the features from the decompiled files feature selection are an important criterion as there is no proper mechanism available for the feature selection process. So for the proper selection of features, the static analysis [5] was performed using the online malware scanners [7]. In the static analysis phase [5], an APK file is given as an input to the website, and the file is uploaded on the website. The file is sent to the server for processing. The server processes the file and returns the results to the website. The user can view the results of the file, and those results are saved to a physical location. All the collected APK malware files were scanned using the online malware scanners, and their reports were collected and analysed. From the analysis of the reports of the online malware scanners, total of 215 features were selected that included various permissions, API calls and Intents. In the feature mining, all the 215 selected features were extracted from the unstructured data collection from the APK files. For the feature mining process, a feature mining script was developed and implemented to extract features from the decompiled APK files. In the feature mining phase [5], the feature mining script looks for a decompiled file in the decompiled files repository and extracts all the 215 features from a decompiled file. A vector for each Android file is generated with extracted features and will be saved in a CSV file. Using this feature mining phase [5], a final data set is generated with a total of 16,300 records in which it includes both malware and benign files [5]. For the evaluation and performance of the generated data set, various supervised machined learning classifiers were implemented in the machine learning phase [6]. In the machine learning phase [6], the features from the generated data set were given as input, and various supervised machine learning classifiers were applied to the features for the classification of the malware and benign files. Different types of supervised classifiers, feature reduction classifiers and ensembling techniques were applied, and the classification result of each classifier is saved into an Excel file. The classification results of each classifier are compared and analysed for better performance and detection of malware. The machine learning phase is explained in our previous paper [6]. The next section describes the overall deployment flow of the AndroHealthCheck system.

## 3 Deployment of AndroHealthCheck

This section discusses the overall deployment flow of the AndroHealthCheck malware detection system. Figure 2 represents the overall deployment flow of AndroHealthCheck. The AndroHealthCheck is divided into four phases such as android file collection, decompilation, feature mining and machine learning. The deployment scenario of the AndroHealthCheck model is discussed according to its phases.

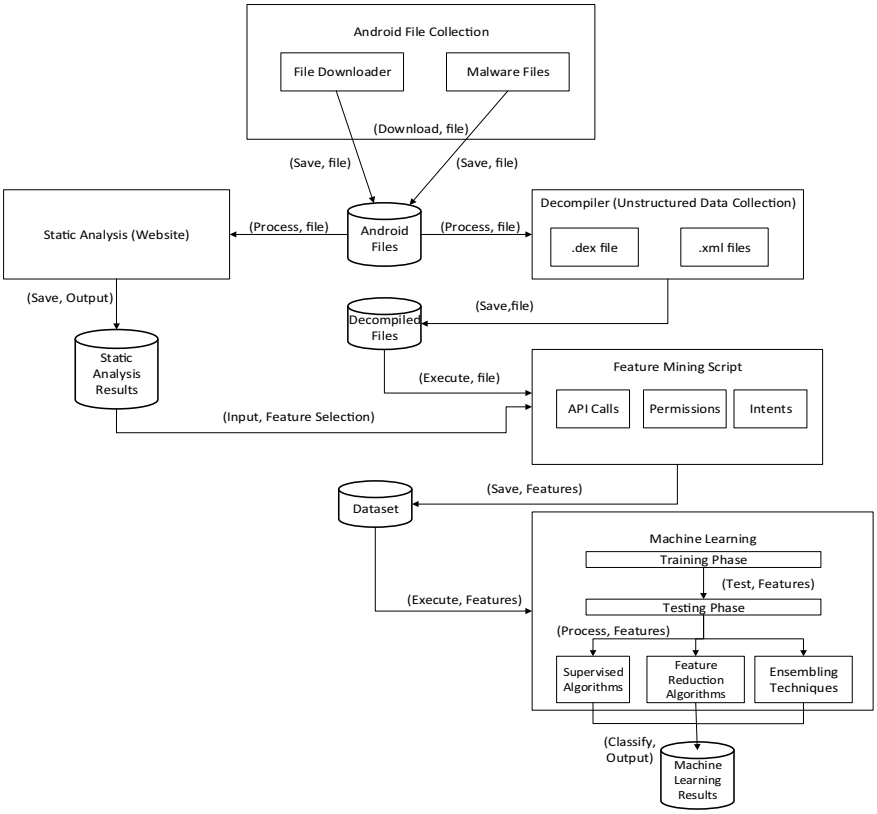



**Fig. 2** Overall deployment flow of AndroHealthCheck

In the android file collection phase, the malware files were downloaded from different android malware projects. A variety of malware files are downloaded from different cloud servers and online sources. For benign files download, we have designed and implemented a crawler/file downloader to automate the process of automatic file download from websites [3]. The crawler is designed using technologies like node package and Cypress framework. The version of Node 10.16.3 is used with Cypress 4.0 along with Chrome browser. The crawler can be deployed on any local machine and executed using Chrome browser. The crawler needs to be initialized with a website URL, and it will fetch and parse a web page by downloading the APK files. The malware files and benign files downloaded are saved into android files repository. In the static analysis phase [5], different online malware scanners were used to process the file and to obtain the results. The downloaded APK file is given as an input to the website, the file processing is done on the cloud servers itself, and the processed zip file is returned to the website and downloaded at a physical location. In the decompilation phase [4], an online decompiler is used for the collection of unstructured data from the APK files. The APK file is given as an input to the website,

and the file is uploaded on the server and is decompiled on the cloud server only. The decompiled file is returned to the website back and saved to a physical location. The decompiled file contains .dex files and .xml files. The .dex files are further processed to extract the Java files.

In the feature mining phase [5], the selected 215 features from analysing the reports of the static analysis phase are extracted from the decompiled APK files. A feature mining script is developed and implemented in Python for extracting the features from the decompiled files. The feature mining script can be deployed on any local machine and executed using Python, and features can be extracted from the decompiled files and saved in a CSV file. The feature mining script mainly extracts API calls, permissions and Intents from the APK files. A vector of extracted features is generated for each Android file and saved in the CSV file. Using the feature mining script a final data set of a total of 16,300 records is generated. For the performance and evaluation of the generated data set of various machines, learning classifiers are used. In the machine learning phase [6], various experiments are carried on the data set by using various machine learning classifiers. The experiments are carried on Intel Core i7-7500U CPU @ 2.90 GHz with 8 GB RAM and Windows 10. The technologies used for experiments are Python and Anaconda Package having a suite of various machine learning libraries for supervised and unsupervised algorithms. For obtaining good results, the data set is divided into a training set and a testing set. From the data set, 75% of data is trained to perform classification and testing on 25% of data. Various machine learning classifiers applied are KNN, random forest, decision tree, linear SVM, logistic regression, Naive Bayes, linear discriminant analysis (LDA), non-negative matrix factorization (NMF), principal component analysis (PCA), bagged decision tree, extra trees and random forest using bagging, gradient boost, CatBoost, AdaBoost, XGBoost, softmax voting and hardmax voting. The classification results of all the machine learning classifiers are stored in an Excel file.

## 4 Conclusion

AndroHealthCheck is a malware detection system to investigate that a file is malware or not by using machine learning methods. A data set is a prerequisite for machine learning models to determine themselves and gain knowledge from the training data for the proper classification of malware and benign files on the testing data. For the investigation of proper malware, a homemade data set with rich variety and with different flavours of malware families was needed. The objective is to create our own data set which can be directly used by researchers for machine learning. A generic malware detection process named AndroHealthCheck is proposed for malware detection using machine learning.

The AndroHealthCheck defines the generic process of data set generation for machine learning and also defines a mechanism for malware detection using machine learning. The architectural and deployment flow of AndroHealthCheck were discussed. In AndroHealthCheck, the design and implementation of a crawler

for automating the process of benign files download were discussed. The unstructured data were collected from the APK files through the decompilation process from all downloaded 15,506 malware and 4000 benign APK files. A proper mechanism is defined for the feature selection process from the APK files through static analysis. The design and implementation of a feature mining script are used for feature mining from unstructured data collection from APK files. A rich data set is generated for machine learning of a total of 16,300 records and 215 features with a huge variety and different flavours of malware files from different families and independent flavours of features. The performance of the generated data set is evaluated with different supervised machine learning classifiers and found that the performance of the CatBoost classifier is highest with 93.15% accuracy and ROC value of 0.91.

**Acknowledgements** We would like to acknowledge our students Ms. Indushree Shetty, Ms. Sabera Kadiwala, Mr. Yash Gajjar, Mr. Ronak Jain, Mr. Vraj Shah and Mr. Akshay Ardeshana of GLSICT, GLS University, who helped us immensely in different phases of AndroHealthCheck a malware detection system for Android.

## References

1. Agrawal P, Trivedi B (2020) Machine learning classifiers for android malware detection. In: 4th International conference on data management, analytics and innovation (ICDMAI). Springer AISC Series, New Delhi, pp 311–322. https://doi.org/10.1007/978-981-15-5616-6_22.ISBN 978-981-15-5616-6
2. Agrawal P, Trivedi B (2019) A survey on android malware and their detection techniques. In: Third international conference on electrical, computer and communication technologies (ICECCT) IEEE, Coimbatore. https://doi.org/10.1109/ICECCT.2019.8868951,E-ISBN 978–1–5386–8158–9
3. Agrawal P, Trivedi B (2020) Automating the process of browsing and downloading APK files as a prerequisite for the malware detection process. Int J Emerg Trends Technol Comput Sci (IJETTCS) 9(2):013–017. ISSN 2278-685
4. Agrawal P, Trivedi B (2020) Unstructured data collection from APK files for malware detection. Int J Comput Appl (IJCA) 176(28):42–45. https://doi.org/10.5120/ijca2020920308.ISBN 973-93-80901-12-5, ISSN 0975 – 8887,
5. Agrawal P, Trivedi B (2020) Feature mining from APK files for malware detection. Int J Appl Inf Syst (IJAIS) 12(32):6–10. https://doi.org/10.5120/ijais2020451874. ISBN 973-93-80975-75-9, ISSN 2249 - 0868
6. Agrawal P, Trivedi B (2020) Evaluating machine learning classifiers to detect android malware. In: IEEE International conference for innovation in technology (INOCON), Bangalore (Paper Selected)
7. Agrawal P, Trivedi B (2019) Analysis of android malware scanning tools. Int J Comput Sci Eng (IJCSE) 7(3):807–810. https://doi.org/10.26438/ijcse/v7i3.807810,E-ISSN 2374–2693

# Use of Machine Learning Services in Cloud


**Chandrashekhar S. Pawar, Amit Ganatra, Amit Nayak, Dipak Ramoliya, and Rajesh Patel**

**Abstract** Machine learning services are the comprehensive description of integrated and semiautomated web devices covering most facilities problems such as preprocessing information, design preparation, and design assessment, with the further forecast. REST APIs can bridge the outcomes of predictions with one's inner IT infrastructure. Like the original SaaS, IaaS, and PaaS cloud delivery models, ML and AI fields cover high-level services to provide infrastructure and platform, exposed as APIs. This article identifying the most used Cloud Technologies for Machine Learning as a Service (MLaaS): Google Cloud AI, Amazon, and Microsoft Azure.

**Keywords** Machine learning (ML) · Machine learning as a service (MLaaS) · Application program interface (API) · Artificial intelligence (AI) · Technology–organization–environment (TOE) framework · Information system (IS)

## 1 Introduction

The development of ML alternatives involves sophisticated, understanding, and costly assets because of this ML was mostly available to big firms that had such capacities. However, it was difficult to harness ML's authority for larger businesses or personal IT experts. One way to address the above-mentioned issues would be to have an ML as a service (MLaaS) capable of providing on-demand computing funds and an obviously specified API for ML procedure. Such a website would allow consumers to concentrate on the issue, and they are attempting to fix rather than the information of execution.

The algorithms of machine learning which are built on deep neural networks (NN) have been widely used in diverse fields. As the use of cloud services grows, MLaaS is accessible, and the training and deployment of these machine learning models are achieved on cloud providers' infrastructure [1].

C. S. Pawar (✉) · A. Ganatra · A. Nayak · D. Ramoliya · R. Patel
Faculty of Technology and Engineering (FTE), Devang Patel Institute of Advance Technology and Research-(DEPSTAR), Charotar University of Science and Technology (CHARUSAT), Changa, Gujarat, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_5

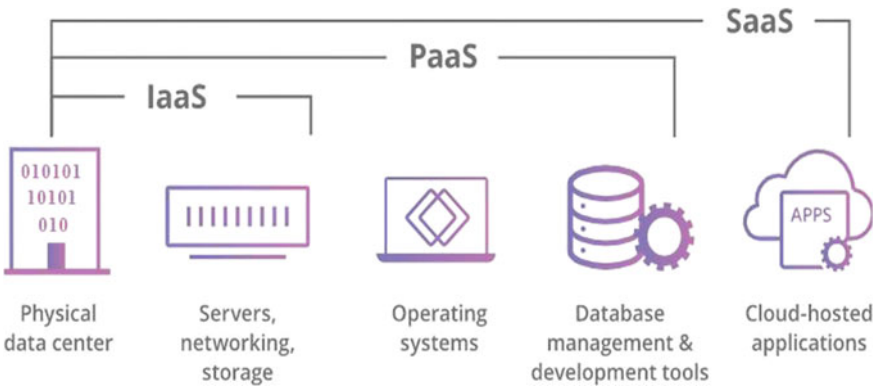



**Fig. 1** Cloud computing service models

Machine training systems are one of the government cloud's highest increasing facilities. ML and AI systems are accessible through various distribution systems such as GPU-based computing, cognitive computing, ML model management, automated machine learning, and ML model serving, and, unlike other cloud-based facilities. Google Cloud AI, Azure machine learning, and Amazon machine learning are major MLaaS cloud amenities that enable quick model instruction and implementation with little or no knowledge in data science [2]. In Fig. 1, the different service models of cloud computing were observed, and Fig. 2 shows how ML cloud services are useful in each service model of cloud computing.

This paper contributes an analysis of Amazon, Google, and Microsoft's primary machine learning as a service application and then compares the machine learning APIs supported by these suppliers. The provided review is not proposed to provide in-depth instructions on what point and how to apply these platforms, but it focuses on what to know before one starts to use machine learning services on the cloud.

## 2 Cognitive Services

Speech recognition, computer vision, and natural language processing services are provided as a set of APIs by cognitive computing [3].

On the daily basis the size of data increase, the use of the services also increases, which force the cloud providers to provide better accuracy for the predictions. With the use of AutoML, a machine learning pipeline with limited computational budgets can be built. AutoML offers an extent territory to temperamental previously trained models vs. training required models from raw information [4].

IBM Watson APIs, Google Cloud AI APIs, Microsoft Cognitive Services, and Amazon AI Services are types of cognitive services. Table 1 gives a comparison of cloud machine learning services.

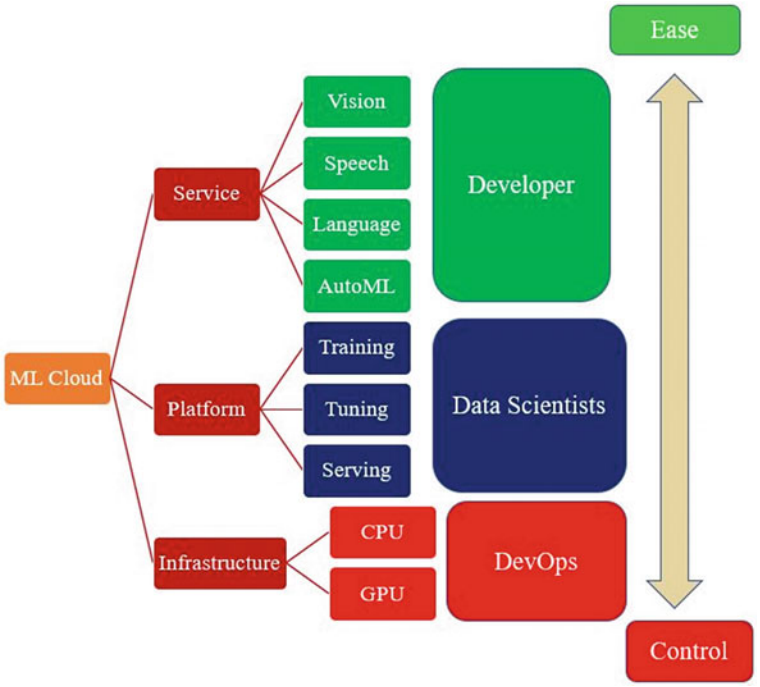



**Fig. 2** ML cloud services

**Table 1** Cloud ML services comparison for custom predictive analytics tasks [5]

| | Amazon | Microsoft | Google | IBM |
|---|---|---|---|---|
| Automated and semiautomated ML services | | | | |
| | Amazon ML | Microsoft azure ML studio | Google prediction API | IBM Watson ML model builder |
| Classification | √ | √ | Deprecated | √ |
| Regression | √ | √ | | √ |
| Clustering | √ | √ | | X |
| Anomaly detection | X | √ | | X |
| Recommendation | X | √ | | X |
| Ranking | X | √ | | X |
| Platforms for custom modeling | | | | |
| Built-in algorithms | √ | X | X | √ |
| Supported frameworks | TensorFlow, MXNet, Keras, Glucon, Pytorch | TensorFlow. Sckit-Learn, Microsoft Cognitive tool-kit, SparkML | TensorFlow, Scikit-Learn, XGBoost, Keras | IBM SPSS, PMML, TensorFlow, Scikit-learn, XGBoost, Spark Mllib |

In [30], through the introduction of cloud-based cognitive radio networks, the user will benefit from a seamless link and wireless Internet access, so the suggested model uses the cloud to control and distribute the space available of license users to the ULU by tracking the spectrum using collective spectrum mapping and the sparse Bayesian algorithm to analyze free space and communicate it to the cloud servers and use a spectrum lessor to assign free spaces to the ULU.

## 3 Machine Learning APIs of Google, Amazon, and Microsoft

MLaaS has become popular during the last few years. The leading Internet companies have deployed their own MLaaS. It offers an easy way to a service provider so that the machine learning (ML) model can be deployed and a quick way for a user/client for making use of the various applications offered by the model [6].

One can provide their own data to these services with trained models and obtain the outcome. APIs do not need proficiency in machine learning. The provided APIs are mainly categorized into three categories as image plus video recognition, text recognition, translation, and textual analysis, and other analysis which includes specific unclassified services.

Table 2 shows the comparison of speech and text processing APIs [5] as below-

Amazon, Microsoft, and Google also provide Video analysis APIs as a service for various analysis purposes. The comparison of these Video Analysis APIs [5] is shown in Table 3.

## 4 ML Services Serve as a Platform

At the point when cognitive APIs miss the mark regarding prerequisites, one can use ML PaaS to construct many tweaked AI models.

For instance, while an intellectual API might almost certainly distinguish the vehicle as a vehicle, it will most likely be unable to order the vehicle-dependent content for advancement and model. Expecting a huge dataset of autos marked with the brand and type, the information science group can depend on ML PaaS to prepare and convey a prototype that is customized for the commercial situation [7].

Like PaaS conveyance archetypal where designers take their code and host it at scale, ML PaaS anticipates that information researchers should bring their individual dataset and code that can prepare a model against custom information. They will be saved from running the register, stockpiling, and systems administration situations to run complex AI occupations [7, 8]. Information researchers are relied upon to make and assess the code with a little dataset in their nearby surroundings before executing it as work in the open cloud stage.

**Table 2** Comparison of speech and text processing APIs [5]

| | Amazon | Microsoft | Google |
|---|---|---|---|
| Speech recognition (speech into text) | √ | √ | √ |
| Text into speech conversion | √ | √ | X |
| Entities extraction | √ | √ | √ |
| Key phrase extraction | √ | √ | √ |
| Language recognition | 100 + languages | 120 languages | 110 + languages |
| Topics extraction | | √ | |
| Spell check | X | √ | X |
| Auto completion | X | √ | X |
| Voice verification | X | √ | X |
| Intention analysis | √ | √ | √ |
| Sentiment analysis | √ | √ | √ |
| Syntax analysis | X | √ | √ |
| Tagging parts of speech | X | √ | √ |
| Filtering inappropriate content | X | √ | √ |
| Low-quality audio handling | √ | X | √ |
| Translation | 6 languages | 60 + languages | 100 + languages |
| Catbot toolset | √ | √ | √ |

ML PaaS expels the grating associated with configuring up and designing information science situations. It gives pre-arranged conditions that can be utilized by information researchers to prepare, configure, and host the model. ML PaaS productively handles the complete execution of an AI model by giving apparatuses from the information willingness stage to demonstrate facilitating.

They accompany mainstream apparatuses, for example, Jupyter Notebooks which are recognizable to the information researchers. ML PaaS handles the unpredictability associated with running the preparation occupations on a bunch of PCs. These are conceptual underpinnings through basic Python or R API for the information researchers [8].

IBM Watson Studio, Amazon SageMaker, Google Cloud ML Engine, and Microsoft Azure ML Services are instances of ML PaaS in the cloud.

In the event that a business needs to bring readiness into AI model improvement and sending, consider ML PaaS. It joins the demonstrated strategy of CI/CD with ML model administration. The ML PaaS is shown in Fig. 3.

**Table 3** Comparison of video analysis APIs [5]

| | Amazon | Microsoft | Google |
|---|---|---|---|
| Object detection | √ | √ | √ |
| Scene detection | √ | √ | √ |
| Activity detection | √ | X | X |
| Facial recognition | √ | √ | X |
| Facial and sentiment analysis | √ | √ | √ |
| Inappropriate content detection | √ | √ | V |
| Celebrity recognition | √ | √ | X |
| Text recognition | √ | √ | X |
| Person tracking on video | √ | √ | X |
| Audio transcription | X | √ | X |
| Speaker indexing | X | √ | X |
| Keyframe extraction | X | √ | X |
| Video translation | X | 9 languages | X |
| Keyword extraction | X | √ | X |
| Annotation | X | √ | X |
| Dominant colors detection | X | X | X |

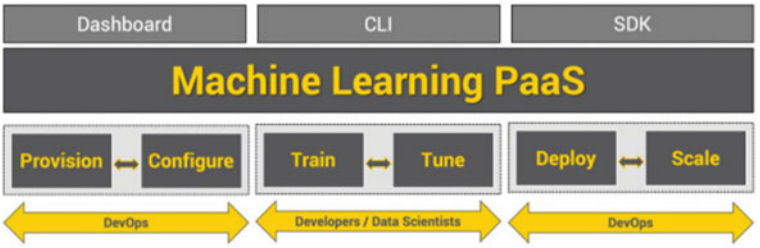



**Fig. 3** Machine learning PaaS

## 5 ML Services Serve as an Infrastructure

Consider ML framework as the infrastructure service of the AI stack. Cloud suppliers recommend crude VMs sponsored by the top of the line CPUs and quickening agents, for example, graphical processing unit (GPU) and field-programmable gate array (FPGA).

Engineers and information researchers that need access to crude process power go to ML foundation. They depend on DevOps groups for arrangement and design required situations. The work process is the same as configuring a testbed for web

or portable application improvement dependent on VMs. From picking the number centers of the CPU to introducing a particular adaptation of Python, DevOps groups claim to start to finish arrangement [9].

For complex profound learning ventures that intensely depend specialty on toolbox and libraries, associations pick ML framework. They deal with the equipment and programming setup which may not be accessible from ML PaaS contributions [10].

The equipment speculations from Microsoft, Google, Amazon, and Facebook are the made ML foundation less expensive and productive. Cloud suppliers are presently configuring equipment as per the requirement that is exceptionally streamlined for running ML remaining burdens in the cloud [11]. Google's TPU and Microsoft's FPGA contributions are instances of custom equipment quickening agents solely implied for ML occupations. At the point when joined with the ongoing processing patterns, for example, Kubernetes, ML framework turns into an appealing decision for endeavors.

Amazon EC2 deep learning AMI sponsored by NVIDIA GPU, Google Cloud TPU, Microsoft Azure Deep Learning VM dependent on NVIDIA GPU, and IBM GPU-based Bare Metal Servers are instances of specialty IaaS for ML.

In Table 4, the past studies were summarized and classified the driving variables (factors) in technical, operational, and environmental contexts, technology adoption.

## 6 ML for Attack Detection and Prevention

Completely decentralized cloud computing and open structure of the Internet and cloud computing. This includes multi-tenancies, multi-domains, and automated multi-user more fragile administrative structures, and sensitive to danger to safety. Missing full oversight of cloud service technology is of great concern to customers. It means the task of detection of intrusion programs and prevention of intrusion programs in protecting cloud computing application information assets [18].

Table 5 [20] represents the paper that suggest deep learning models and big data analytics for collecting and analyzing information.

Intrusion prevention systems (IPS) will respond to any anomalies found. IPS shuts down access of the server infected or the device program (it/IP). The address the detected traffic originates from is blocked. Unlike IDS and IPS, intrusion can be a host or network-based Prevention programs according to whether or not they are protects a host or operate at the network level. These methods of antiviruses are used for the prevention [19].

Today's IDPS required improved aspect architecture of speed and quantity of data required Precise. Cloud services come with elastic tools network bandwidth, data, power transmitter, virtualization, high-quality delivery, and accordingly, they form the basis for the analysis of big data and the creation of through cloud services, security services.

**Table 4** Study using the structure for the TOE in IS disciplines [12]

| Refs no | Study | Technological context | Organizational context | Environmental context |
|---|---|---|---|---|
| [13] | Implementation of cloud computing | Intricacy compatibility of relative advantage | Top management assistance firm scale preparation for technology | Strong competition, the pressure of trading partners |
| [14] | RFID adoption in Chinese companies | Gain usability expense use of IS firm size | IS unit practitioner's Top management service | Consumer competitor RFID supplier government |
| [15] | In the Malaysian tourism market, e-commerce use and business performance | Competence of technology | Firm size business scale Web investment technology Management beliefs | Regulatory support strength of the pressure |
| [16] | Innovation as-simulation, a view of technological diffusion | Integration of Technology Readiness | Size global barriers to management | Regulatory environment: rivalry level |
| [17] | Effective B2B e-commerce (adoption and diffusion of e-commerce) | Service discontinuity Compatibility incorporation Benefits of emerging technologies EDI precision of assets | IT policymakers readiness-knowledge administrative framework | Competitive environment market partner partnership dynamics of industry external capital support for industry institutional factors |

## 7 Conclusion

Machine learning and artificial intelligence can be expensive—skills and resources can cost a lot. For that reason, MLaaS goes to be a massively influential development inside the cloud. The ranges of services are offered from AWS, Azure, and GCP. It is really the ease and convenience that is most remarkable. To enhance the business processes and operations, client communications, and in general business policy, these services help due to their effortless configuration and execution of ML algorithms. So, this review work can help the readers to check and decide which ML service is available in which cloud model.

**Table 5** Summary of paper which analyze the attacks using deep learning and big data analytics

| Ref No. | Dataset | Algorithm and technology | Research field |
|---|---|---|---|
| [21] | User dataset | MOA, NeuralNetwork, Filing Storm, Spark Storm, Hadoop | New techniques for NIDS in cloud |
| [22] | User dataset | Hadoop, Ganglia and Nagios1 | Architecture for anomaly detection |
| [23] | KDD-99 | Restricted Boltzmann Machine (RBM), Logistic Regression | Real-time novel attacks detection |
| [24] | – | Predictive performance anomaly prevention | Comprehensive analytics in business, Service-Level Agreement (SLA) |
| [25] | – - | Twitter: R studio, Hadoop | Limits of relational databases |
| [26] | User dataset | HAMR, next-generation in-memory MapReduce engine | Detection of anomalies on a distributed architecture |
| [27] | User dataset | Hadoop, decision tree | Detection of anomalies on the Big Data platform |
| [28] | User dataset | Hadoop, eagle, density estimation, PCA | User behavior analysis |
| [29] | User dataset | Net sniffer NetL, process log, HDFS, k-means | Detection of anomalies over events from various logs |

## References


1. Hesamifard E et al (2018) Privacy-preserving machine learning as a service. In: Proceedings of privacy enhancing technologies, pp 123–142
2. Baldominos A, Albacete E, Saez Y, Isasi P (2014)A scalable machine learning online service for big data real-time analysis. In: IEEE symposium on computational intelligence in big data (CIBD). IEEE, pp 1–8
3. Ghoting et al (2011) SystemML: declarative machine learning on MapReduce. In: Proceedings of the 2011 IEEE 27th international conference on data engineering, ICDE 11. Washington, DC, USA, pp 231–242
4. He X, Zhao K, Chu X (2019) AutoML: a survey of the state-of-the-art. arXiv preprint arXiv: 1908.00709
5. Comparing Machine Learning as a Service: Amazon, Microsoft Azure, Google Cloud AI, IBM Watson, https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-cloud-ai-ibm-watson. Last accessed 20 Dec 2019
6. Pop D, Iuhasz G (2011) Survey of machine learning tools and libraries. Institute e-Austria Timisoara Technical Report
7. Alpaydin (2014) Introduction to machine learning. MIT press, India
8. Mohri M, Rostamizadeh A, Talwalkar A (2012) Foundations of machine learning. The MIT Press
9. Sagha H, Bayati H, Millán JDR, and havarriaga R (2013) On-line anomaly detection and resilience in classifier ensembles. Pattern Recogn Lett Elsevier Science Inc. 34:1916–1927
10. Stephen F. Elston: Data Science in the Cloud with Microsoft Azure Machine Learning and R. O'Reilly
11. Low Y, Gonzalez J, Kyrola A, Bickson D, Guestrin C, Hellerstein JM (2012) Distributed GraphLab: a framework for machine learning and data mining in the cloud. In: Proceedings of the VLDB endowment, vol 5, no 8, Istanbul, Turkey

12. Saedi A, Iahad NA (2013) Future research on cloud computing adoption by small and medium-sized enterprises: a critical analysis of relevant theories. Int J Actor-Netw Theor Technol Innov (IJANTTI) 5(2):1–6
13. Low C, Chen Y, Wu M (2011) Understanding the determinants of cloud computing adoption. Ind Manag Data Syst 111:1006–1023. https://doi.org/10.1108/02635571111161262
14. Li J, Wang Y-F, Zhang Z-M, Chu C-H (2010) Investigating acceptance of RFID in Chinese firms: the technology-organization-environment framework. In: Program for the IEEE international conference on RFID-technology and applications, Guangzhou, China
15. Salwani MI, Marthandan G, Norzaidi MD, Chong SC (2009) E-commerce usage and business performance in the Malaysian tourism sector: empirical analysis. Inf Manag Comput Secur 17:166–185. https://doi.org/10.1108/09685220910964027
16. Zhu K, Kraemer KL, Xu S (2006) The of innovation assimilation by firms in different countries: A technology diffusion perspective. Manage Sci 52:1557–1576. https://doi.org/10.1287/mnsc.1050.0487
17. Robertson RA (2005) A framework of critical drivers in successful business-to-business e-commerce. In: Proceedings of the IEEE Southeast conference, pp 378–38, 8–10 Apr 2005
18. Taghavi M, Bakhtiyari K, Júnior JC, Patel A (2013) An intrusion detection and prevention system in cloud computing: a systematic review. J Network Comput Appl 36(1):25–41
19. Scarfone K, Mell P (2007) Guide to intrusion detection and prevention systems (IDPS). National Institute of Standards and Technology Special Publication 800-94
20. Lidong W, Randy J (2017) Big data analytics for network intrusion detection: a survey. Int J Netw Commun 2017:24–31
21. Son S, Gil M-S, Moon Y-S (2017) Anomaly detection for big log data using a hadoop ecosystem. In: IEEE international conference on big data and smart computing (BigComp), Jeju, South Korea
22. Alrawashdeh K, Purdy C (2016) Toward an online anomaly intrusion detection system based on deep learning. In: 15th IEEE international conference on machine learning and applications, Anaheim, CA, USA
23. Ramamohanarao K, Leckie C, Buyya R, Calheiros RN, Dastjerdi AV, Versteeg S (2015) Big data analytics-enhanced cloud computing: challenges, architectural elements, and future directions. In: IEEE 21st international conference on parallel and distributed systems, Melbourne, VIC, Australia
24. Disha DN, Sowmya BJ, Chetan, Seema S (2016) An efficient framework of data mining and its analytics on massive streams of big data repositories. In: Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). IEEE, Mangalore, India
25. Janeja VP, Azari A, Namayanja JM, Heilig B (2014) B-dids: mining anomalies in a big-distributed intrusion detection system. In: IEEE international conference on big data. Washington, DC, USA
26. Gupta P, Stewart C (2016) Early work on characterizing performance anomalies in hadoop. In: IEEE international conference on autonomic computing (ICAC), Wurzburg, Germany
27. Gupta C, Sinha R, Zhang Y (2015) Eagle: user profile-based anomaly detection for securing hadoop clusters. In: IEEE international conference on big data (Big Data), Santa Clara, CA, USA
28. Razaq A, Tianfield H, Barrie P (2016) A big data analytics based approach to anomaly detection (BDCAT). In: IEEE/ACM 3rd international conference, Shanghai, China
29. Avdagic I, Hajdarevic K (2017) Survey on machine learning algorithms as cloud service for CIDPS. In: 25th Telecommunication forum (TELFOR), Belgrade, Serbia
30. Bindhu V (2020) Constraints mitigation in cognitive radio networks using cloud computing. J Trends Comput Sci Smart Technol 2(1):1–14

# An Experimental Analysis on Selfish Node Detection Techniques for MANET Based on MSD and MBD-SNDT


**V. Ramesh and C. Suresh Kumar**

**Abstract** Mobile ad hoc network (MANET) is a network that permits mobile servers and customers to convey without fixed infrastructure. MANET is a promptly developing region of investigation as it employs a group of uses. To encourage effective information access and update data sets are conveyed on MANET. Since information openness is influenced by the portability and power imperatives of the servers and customers, the information in MANET is reproduced. As in ad hoc network, since portable host moves unreservedly network panel happens habitually in this way information accessibility is decreased bringing about execution debasement. The majority of the replication strategies expect that each node advances each packet provided to it and combine completely as far as sharing the memory space. A portion of the nodes may go about as selfish nodes which may selfishly conclude and participate incompletely with all the different nodes. In this manner, selfish conduct of nodes could lessen the general information openness in the network. At first, a design model of a MANET is built and the correspondence between the versatile is started. The packet drop can occur in MANET because of the selfish node or network clog. In this paper, MSD-SNDT and MBD-SNDT method is proposed to recognize the selfish nodes effectively in MANET. The reproduction study shows that the proposed MSD-SNDT and MBD-SNDT strategy improves the selfish node detection ratio, packet delivery proportion (PDP), and normal packet drop proportion.

**Keywords** Selfish nodes · Selfish node detection technique · Throughput · Packet delivery rate · Modified Skellam distribution (MSD) · Modified Bates distribution (MBD) · Constant bit rate (CBR)

V. Ramesh
Department of Computer Science, Bharathiar University, Coimbatore, Tamilnadu, India

C. Suresh Kumar (✉)
DR. Nagarathinam College of Engineering, Namakkal, Tamilnadu, India
e-mail: principaldrsureshkumar@gmail.com



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_7

## 1 Introduction

In a mobile ad hoc network (MANET), the constancy of the portable nodes toward pack sending is made plans to be a basic substance for ensuring dominating bundle sending rate. This unfaltering nature of mobile nodes is influenced by their prideful lead toward data sending development to remain dynamic in the framework. A gigantic number of fundamental applications, exceptionally selected express the control over the frameworks and achieved fewer troubles and stresses concerning assurance and security of the data toward the coordinating information. The framework with intrinsic exceptional geography is crucial clarification of security-related concerns. Establishing fewer frameworks organizes the absence of united control impacts the establishment and restoration of security-related information. The center points present in the framework bit by bit extending the coordinating overhead, since the groups are separated by the flexible centers while sending them.

Skellam scattering is considered as the most serious scattering that models the boundaries used for intolerant center point. It is a recognizable proof for the discrete sporadic variable for achieving common revelation. Moreover, it is recognized for assessing the particular estimation about the negative impact constrained by the selfish nodes under coordinating. Along these lines, a biased node revelation and disengagement approach that depends upon Skellam scattering is essential for the exact improvement in the framework to extent the package during transport and throughput.

This proposed MSD-SNDT and MBD-SNDT utilize the key preferences of Skellam and Bates distribution for ensuring potential selfish acknowledgment measure.

## 2 Literature Review

From the outset, a selfish node id scheme using essential level and extent of node correspondence was propounded for perceiving the malignant intolerant development of the portable nodes in the framework [1]. The imperative level and extent of node correspondence are based on the narcissistic NODES area approach. It was assessed to improve the movement of recognizable proof with the ultimate objective that the selfish motivation behind the mobile nodes is prevented as far as possible.

Network monitor canine [2] approach and a deliberate model that evaluates season of distinguishing the extremist nodes, they have extended the work incase pretentious node assembles a mean-max gauge for achievable calculation. The display of the on-demand multi-way controls [3] in portable exceptionally designated frameworks, and the proposed structure is conceivable and versatile to find the most concise method to send the data groups in a secured strategy with malicious node distinguishing proof. The pleasing distinguishing proof attack is watching [4] all the new segment nodes in sort out, accommodating nodes are passed on between the neighbor nodes two

successive nodes in the course, and if attacker nodes are perceived, by then reversing to the all-nodes through mix places.

An enthusiastic pleasing trust [5] building up a plan for reaching on the groups securely and constantly in multi-ricochet courses, in the arrangement choosing the trust for each center. In the indirect node trust, it is happened by MAC layer and reused node is by recommendation of the neighboring nodes to perceive and send packages; this is incredible on the counterfeit information of noxious nodes. Moreover, Distributed Detection Of Selfish Nodes Utilizing Dynamic Associativity (DDSN-DA) was proposed for diminishing the degree of false revelation in the biased centers of the framework [7]. The degree of trust and acknowledgment supported by this DDSN-DA plot was turned being expanded during the route toward arranging mobile nodes into extremist and reliable. Finally, An exponential reliability factor-based selfish node detection technique (ERF-SNDT) was proposed for likely id and controls the selfish nodes in the framework [8]. This ERF-SNDT approach was set to diminish the group drop, essentialness usages, bundle inaction, and hard and fast overhead to the great level that appeared differently with the DDSN-DA and SRA-FSND approaches added to potential biased center point acknowledgment measure. The movement of a false certain movement of this proposed ERF-SNDT approach was set out to be generally extraordinary with diminished overhead in figuring and correspondence of the framework.

Finally, the semi-Markov process mechanism utilizing selfish node detection technique (SMPM-SNDT) was proposed for productive gauging in malignant development subject to the current status of bundle sending rate [6]. The SMPM-SNDT was assessed for maintaining unparalleled execution to the extent of improved throughput, diminished total overhead, and control overhead [9].

## 3 Related Work

The authors [10] overviewed the problems and pointers to mild execution corruption and network packing in MANET's. A large portion of the tries to alleviate selfish behavior might be ordered into the incentive-based mechanism, and reputation-primarily based on mechanism and various additives. The authors [11] format self-designing because of the capability to adjust automated associate and powerfully to herbal adjustments. The difficulty of animating collaboration in self-arranging transportable unintentional companies for non-military network packages is self-tended to [12]. This methodology is utilized to regulate secure system module alluded to as a safety module in every mobile node. A credit-based complete convention is applied to animate the participation among flexible nodes in packet sending. This is a good way to upgrade the employer execution through relieving the miserly conduct [13]. The authors [14] organized a decent, sensible and at ease collaboration in motivating the pressure of the machine for multi-jump remote businesses. To parsimony

assaults and invigorate hub collaboration is enhanced to upgrade the business enterprise execution and reasonableness. It is performed by means of charging each of the appropriate and target nodes of the correspondence.

Selfish node detection technique exploitation organization head became predicted call and dependableness of flexible mobile nodes inside the network [15]. This cluster head-based call turns the difficulty into the firm level of directing overhead and power utilizations. This contrasted the canine-based totally agreeable to the contact technique. The distinguishing evidence attack is attentive [4], and the all of the new segment middle point in arrange, precious middle point is arranged in the rectangular degree which is passed among the neighbor hubs and resulting hubs in course. Within this event, the attacker facilities are apprehend, and round then add weights to the all middle factors via combo locations. SHRCDM changed into preparations for encouraging crucial participation among the mobile nodes [8]. This SHRCDM engine aided better segregation of mobile nodes to agreeable and stingy nodes for main execution development inside the network.

In this nearby aspect-based totally miserly node identity technique includes a number of the chose nodes utilized in rectangular measure for investigating the features of mobile nodes to look the extent of deliberate behavior ascribed by them toward the network. A gradable area conscious Hash Table-Based totally Selfish Node Detection Mechanism (HLAHT-SNDM) turned into contributed through ceaseless notion [16]. Finally, the Hash Table-Based totally Selfish Node Detection Technique (HLAHT-SNDT) changed into made preparations for effective predicting in malevolent action upheld the contemporary remaining of packet sending charge [9]. The SMPM-SNDT changed into measurable for actualizing universal execution as a long way as elevated outturn, reduced the overhead and executives overhead. The identification pace of this system head-based totally calls technique to become conjointly resolved to the maximum under any assortment of selfish nodes inside the network.

## 4 Modified Skellam Distribution -Based Selfish Node Detection Technique (MSD-SNDT)

The MSD-SNDT is created in this examination for identifying Selfish Nodes.

MSD-SNDT follows three stages:

(A) Computing mean packet deviation
(B) Calculation of variance and standard deviation for computing MSD
(C) Detection of such nodes.

(A) ***Computing mean packet deviation***

The deviation in the number of packets received to the number of the packet forwarded by each mobile node to their neighbors as recommended through neighbor-based interaction in each session '$c$' is

$$\text{DEVIATION}_{\text{PACKET}(c)} = \text{PR}_{(c)} - \text{PF}_{(c)} \tag{1}$$

$\text{PF}_{(1)}$, $\text{PF}_{(2)}$, …., $\text{PF}_{(s)}$ and $\text{PR}_{(1)}$, $\text{PR}_{(2)}$, …, $\text{PR}_{(s)}$ define packet forwarded and packet received, respectively, by each of its neighbors in 's' sessions.

Packet forwarding capability identified by their neighbor is

$$P_{\text{PFC}(C)} - \frac{\text{PF}_{(C)}}{\text{PR}_{(C)}} \tag{2}$$

The mean packet deviation is given by

$$\text{MDEV}_{\text{PACKET}(C)} = \sum_{c=1}^{s} \frac{\text{DEVIATION}_{\text{PACKET}(C)}}{S} \tag{3}$$

(B) ***Calculation of variance and standard deviation for computing MSD***

$$\text{STD}_{\text{DETECT}} = P_{\text{PFC}(s)} * (1 - P_{\text{PFC}(s)}) \tag{4}$$

$$\text{VARIANCE}_{\text{DETECT}} = \sum_{c=1}^{s} (\text{MDEV}_{\text{PACKET}(c)} - \text{DEVIATION}_{\text{PACKET}(c)})2 \tag{5}$$

(C) ***Detection of such nodes***

Then MSD computed based on (4) and (5) is

$$\text{MSD}_{\text{DETECT}} = \frac{s}{s-1}\left(1 - \frac{\sum_{c=1}^{s} \text{STD}_{\text{DETECT}}}{\text{VARIANCE}_{\text{DETECT}}}\right) \tag{6}$$

## 5 Detection and Isolation of Selfish Nodes Misbehavior Utilizing Computed MSD

The mobile nodes found with MSD esteem under 0.35 are distinguished as selfish nodes and isolated.

### *5.1 The Proposed Algorithm—MSD-SDNT*

The accompanying algorithm represents the means engaged with distinguishing selfish nodes utilizing MSD

### 5.1.1 The Steps for Proposed Algorithm MSD-SNDT

1. Let N be the number of Nodes.
2. Let GN be Group Node (GN), in which SN is Source Node and DN is Destination Node.
3. Set of nodes in the routing path can be set up by sending 'RREQ' message by the NS to all different nodes in the network
4. Mobile node reacts to the source node by 'RREP'.
5. Let this algorithm STEP (6–14) be executed for a node say, k, which has a place with the rundown GN that utilizes 't' number of sessions for transmission.
6. For each node 'k' of GN in the routing path.
7. Determine deviation N utilizing condition 1.
8. Compute packet forwarding capacity utilizing Eq. 2.
9. Calculate mean deviation utilizing Eq. 3.
10. Using Eqs. 4 and 5 decide STD and variance separately
11. Compute MSD utilizing Eq. 6.
12. if (MSD(k) < 0.35) at that point
13. node k is selfish node misbehavior bargained
14. Call Selfish_Node_-Mitigation (k)
15. Else
16. node k is reliable.
17. End if
18. End for
19. End for

## 6 Experimental Results and Discussions of MSD-SNDT

The predominant function of the proposed MSD - SNDT plot is researched the simulation experiments utilizing ns-2.31. The simulation time for the execution is 100 s with the CBR traffic pattern of data. The quantity of mobile nodes in the network is 100 disseminated randomly with the size bytes of 512 packets.

Figure 1 models the throughput of the proposed MBD-SNDT plot examined under a substitute number of mobile nodes in the framework. The proposed MBD-SNDT is found to increase the throughput to a most outrageous level of 11, 13, and 18% better than the ERF-SNDT, DDSN-DA, and SRA-FSND approaches.

Figure 2 portrays the control overhead of the MBD-SNDT plot researched under a substitute number of portable nodes in the framework. The proposed MBD-SNDT is shown to restrict the control overhead to a most extraordinary level of 10, 14, and 18% better than taking a gander at ERF-SNDT, DDSN-DA and SRA-FSND

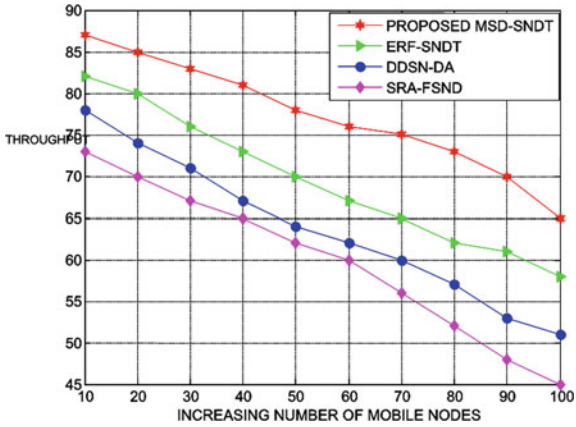



**Fig. 1** Performance of MSD-SNDT-throughput-different mobile nodes

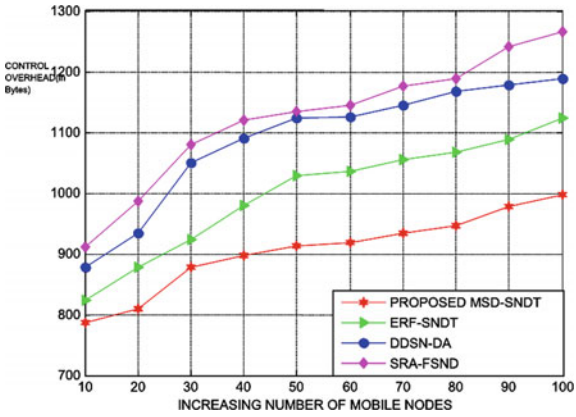



**Fig. 2** Performance of MSD-SNDT-control overhead-mobile nodes

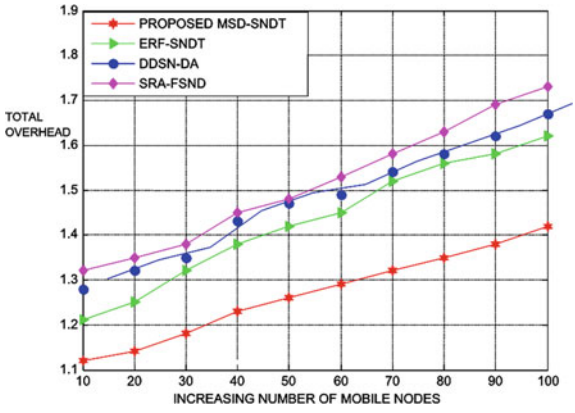



**Fig. 3** Performance of MSD-SNDT-total overhead-different nodes

Figure 3 highlights the total overhead of the proposed MBD-SNDT plot researched under a substitute number of portable nodes in the framework. The proposed MBD-SNDT is set out to diminish total overhead to the biggest level of 9, 12, and 18% better than the idea about ERF-SNDT, DDSN-DA, and SRA-FSND approaches.

Figure 4 presents the plots in the group lethargy of the proposed MBD-SNDT plan investigated under a substitute number of mobile nodes in the framework. The proposed MBD-SNDT is set out to check the pack inertness to a great level of 12, 18%, and 21 better than the dissected ERF-SNDT, DDSN-DA, and SRA-FSND.

Figure 5 highlights the throughput of the proposed MBD-SNDT plan explored under a substitute number of pompous nodes in the framework. The proposed MBD-SNDT is inferred to grow the throughput to a most extraordinary level of 11, 15, and 19% better than they took a gander at ERF-SNDT, DDSN-DA, and SRA-FSND approaches.

Additionally, Fig. 6 models control overhead of the proposed MBD-SNDT plot examined under a substitute number of prideful NODES in the framework. The

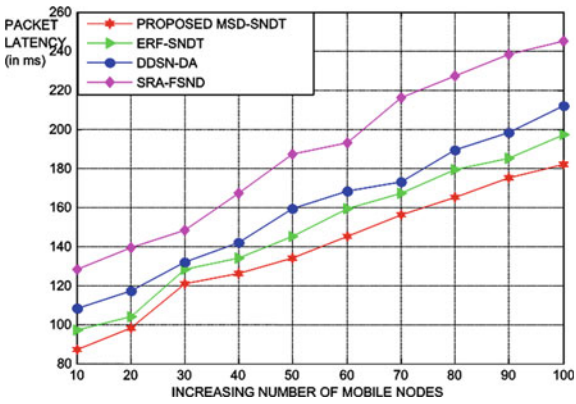



**Fig. 4** Performance of MSD-SNDT-packet latency-different mobile nodes

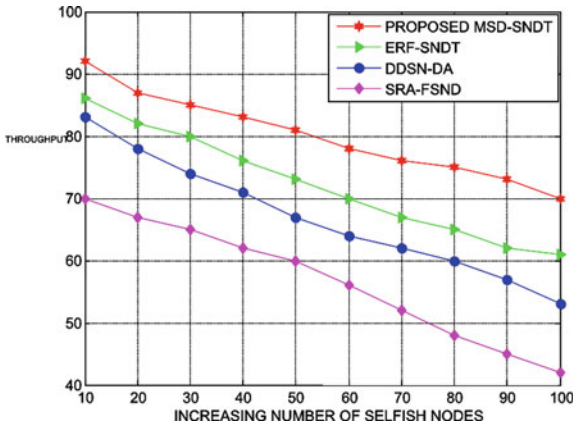



**Fig. 5** Performance of MSD-SNDT-throughput-different selfish nodes

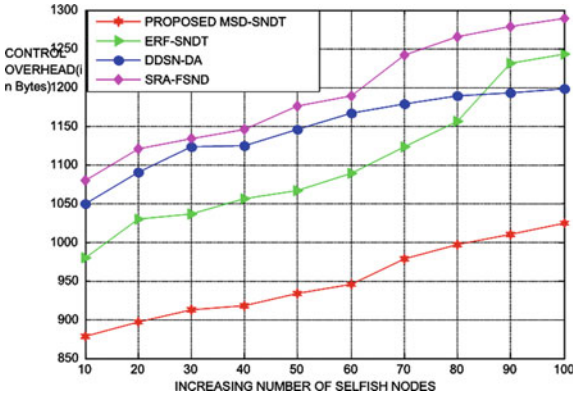



**Fig. 6** Performance of MSD-SNDT-control overhead-selfish nodes

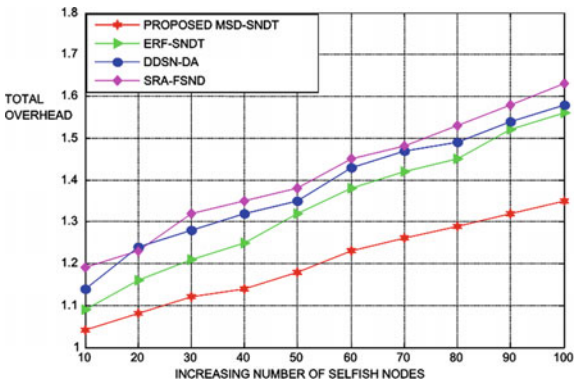



**Fig. 7** Performance of MSD-SNDT-total overhead-different selfish nodes

proposed MBD-SNDT is shown to restrict the control overhead to the biggest level of 11, 14, and 16% better than the dissected ERF-SNDT, DDSN-DA and SRA-FSND approaches.

Figure 7 depicts the proposed MBD-SNDT is set out to decrease full-scale overhead to a most outrageous level of 9%, 12%, and 16% better than they took a gander at ERF-SNDT, DDSN-DA, and SRA-FSND approaches.

Figure 8 uncovers the proposed MBD-SNDT is set out to restrict the packet inactivity to a critical level of 12, 15 and 18% better than the examined ERF-SNDT, DDSN-DA, and SRA-FSND approaches.

## 7 Modified Bates Distribution—Based Selfish Node Detection Technique (MBD-SNDT)

The projected MBD-SNDT, the detection, and separation of selfish nodes include

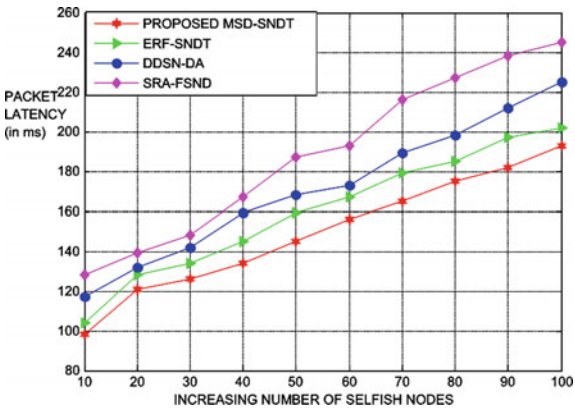



**Fig. 8** Performance of MSD-SNDT-packet latency-different selfish nodes

1. Intention of mean packet deviation,
2. Assessment of variance and standard deviation for computing MBD and
3. Detection and separation of selfish node

(a) **Mean Packet Drop**

If the number of packets forwarded and obtained by using every node '*i*' is PF(1), PF(2), …, PF(s) and PR(1), PR(2), …, PR(s), respectively, as monitored by every one of its peers in '*s*' sessions. The range of packets dropped via every mobile node as monitored via their neighbors in every consultation '*k*' is

$$\mathrm{DROP}_{\mathrm{PACKET}(k)} = \mathrm{PR}_{(k)} - \mathrm{PF}_{(k)} \tag{7}$$

The mean packet drop is

$$\mathrm{MDROP}_{\mathrm{PACKET}(k)} = \sum_{k=1}^{s} \frac{\mathrm{DROP}_{\mathrm{PACKET}(k)}}{S} \tag{8}$$

(b) **Total variance**

The total variance is

$$T - \mathrm{VAR}_{\mathrm{DETECT}} = \sum_{c=1}^{s} \frac{(\mathrm{PR}(C) - \mathrm{MDROP}_{\mathrm{PACKET}(C)})}{S} \tag{9}$$

The MBD computed based on (9) and (10) is

$$(\mathrm{BDITF})_{\mathrm{DETECT}} = \frac{S}{S-1}\left(1 - \frac{\sum_{k=1}^{s} \mathrm{DROP}_{\mathrm{PACKET}(k)} * \mathrm{PR}_{(k)}}{T - \mathrm{VAR}_{\mathrm{DETECT}}}\right) \tag{10}$$

The mobile nodes are recognized by way of MBD less than 0.35 (received from simulation) are detected as selfish node misbehavior.

## *7.1 MBD-SNDT Algorithm*

The subsequent algorithm illustrates the stairs involved in detecting selfish node misbehavior the use of MBD and keeping apart them from the multicasting activity.

| | |
|---|---|
| 1. | Let N be a number of nodes. |
| 2. | GN – Group of nodes of the routing path, SN (the source node) and DN (the destination node), respectively. |
| 3. | The cell node which is prepared for statistics transmission acknowledges SN through 'RREP' message. |
| 4. | Allow this set of rules step (5–12) to be accomplished for a node 'n' wide variety of sessions for transmission. |
| 5. | For every node 'u' of GN in the routing course |
| 6. | Estimate $DROP_{PACKET(k)} = PR_{(k)} - PF_{(k)}$ |
| 7. | Compute $MDROP_{PACKET(k)} = \sum_{k=1}^{S} \frac{DROP_{PACKET(k)}}{S}$ |
| 8. | Calculate $T - VAR_{DETECT} = \sum_{c=1}^{S} \frac{(PR_{(C)} - MDROP_{PACKET(C)})}{S}$ |
| 9. | Estimate $(BDITF)_{DETECT} = \frac{S}{S-1}\left(1 - \frac{\sum_{k=1}^{S} DROP_{PACKET(k)} * PR_{(k)}}{T-VAR_{DETECT}}\right)$ |
| 10. | If (MBD(u) < 0.35) then |
| 11. | Node u is selfish node misbehavior compromised |
| 12. | Name selfish_node_attack-mitigation (u) |
| 13. | Else |
| 14. | Node u is reliable |
| 15. | End if |
| 16. | End for |
| 17. | End |

## 8 Experimental Results and Discussions of MBD-SNDT

The strength of the proposed MBD-SNDT is explored dependent on duplicate tests led utilizing ns-2.33. The capability of the proposed MBD-SNDT approach is contrasted and the modern SMPM-SNDT, HLAHT-SDNM, and SHRCDM approaches utilizing the packet conveyance share, throughput, entire overhead and also utilizes electricity underneath the expanding variety of flexible hubs and narrow-minded hubs. The facts visitors layout utilized for utilization are constant bit rate (CBR) information traffic. The simulation time applied for the proposed MBD-SNDT is 300 s with an offseason of 20 s.

Figures 9 and 10 model the capability of the proposed MBD-SNDT method utilizing package conveyance proportion and throughput researched under the expanding pace of mobile nodes. The packet conveys proportion of the proposed MBD-SNDT technique which is affirmed to be the main via 6%, 10%, 13% contrasted, and the cutting-edge SMPM-SNDT, HLAHT-SDNM and SHRCDM processes. Consequently, the throughput of the proposed MBD-SNDT approach

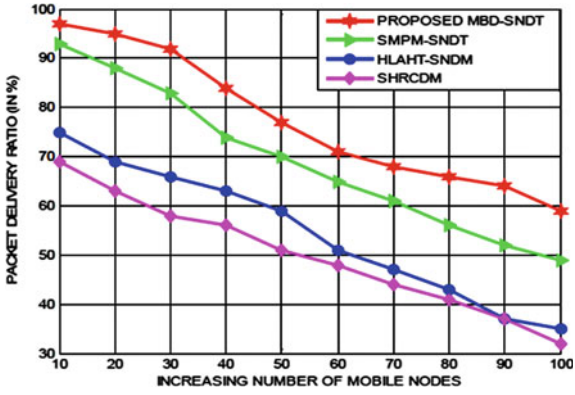



**Fig. 9** Performance of MBD-SNDT utilizing package delivery ratio under increasing mobile nodes

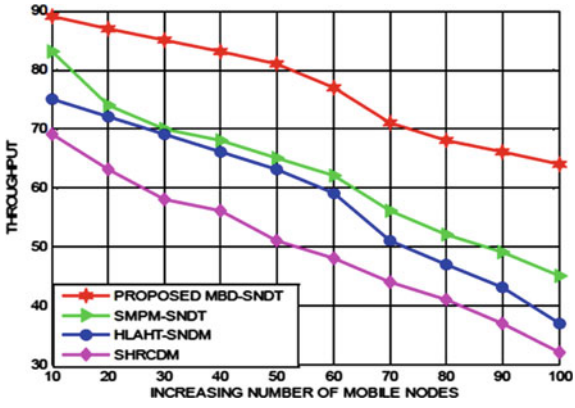



**Fig. 10** Performance of MBD-SNDT making use of throughput underneath increasing mobile nodes

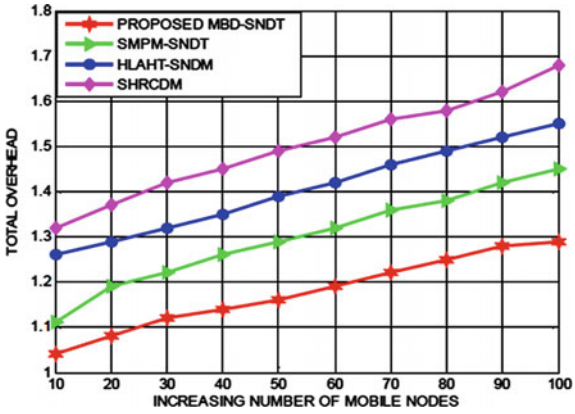



**Fig. 11** Performance of MBD-SNDT making use of packet delivery ratio below expanding mobile nodes

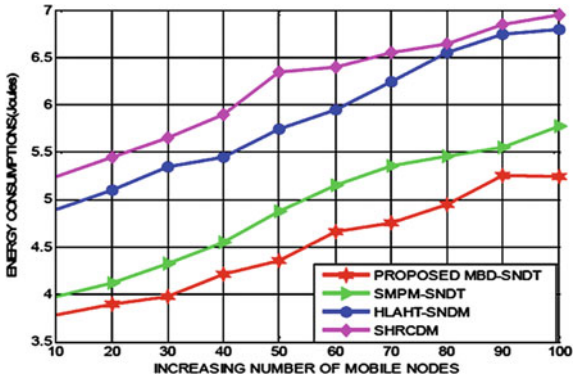



**Fig. 12** Overall performance of MBD-SNDT utilizing energy utilization under increasing mobile nodes

is resolved to be surprising via 9, 11, 14% contrasted, and the cutting-edge SMPM-SNDT, HLAHT-SDNM, and SHRCDM processes.

Figures 11 and 12 exhibit the PROPOSED MBD-SNDT method using all-out overhead and strength utilizations examined under the increasing tempo of mobile nodes. The entire overhead of the proposed is basically restrained through 10%, 13%, and 15% extraordinary with the contemporary SMPM-SNDT, HLAHT-SDNM, and SHRCDM approaches. Moreover, the power utilization of the proposed technique is resolved to be enormously decreased via 8, 10, and 13% contrasted, and the cutting-edge SMPM-SNDT, HLAHT-SDNM, and SHRCDM procedures.

Figures 13 and 14 measure the capability of the proposed SDITF-SNDT method using packet conveyance share and throughput examined below the expanding tempo of selfish nodes. The packet deal conveyance share of the proposed technique underneath expanding selfish nodes is affirmed to improve 9%, 13%, and 16% impressive to the benchmarked SMPM-SNDT, HLAHT-SDNM, and SHRCDM procedures. Basically, the throughput of the proposed approach under increasing selfish nodes

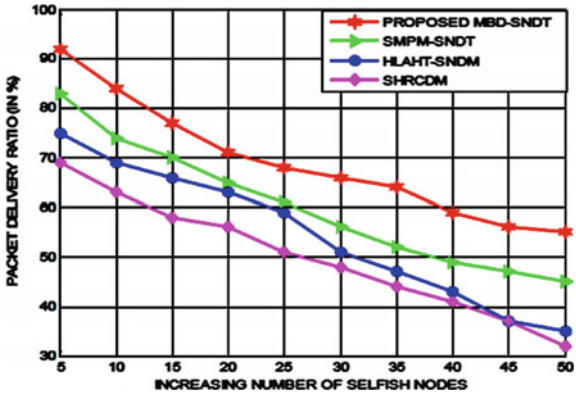



**Fig. 13** Overall performance of MBD-SNDT utilizing packet delivery ratio share under increasing selfish nodes

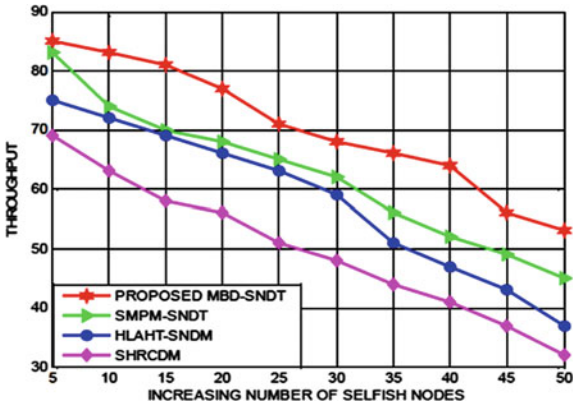



**Fig. 14** Overall performance of MBD-SNDT utilizing energy utilization under expanding mobile nodes

is resolved to be extremely good via 10, 12, and 15% contrasted and the current SMPM-SNDT, HLAHT-SDNM, and SHRCDM tactics.

Figures 15 and 16 show capabilities the criticalness of the proposed MBD-SNDT approach utilizing absolute overhead and strength utilizations researched underneath increasing tempo of slender minded hubs. Absolutely the overhead of the proposed technique is resolved to be essentially limited using 10%, 13%, and 16% immediate to the modern-day SMPM-SNDT, HLAHT-SDNM, and SHRCDM tactics. Additionally, the strength utilizations of the proposed approach are resolved to be magnificently diminished by 7, 9, and 12% contrasted and the current SMPM-SNDT, HLAHT-SDNM, and SHRCDM approaches.

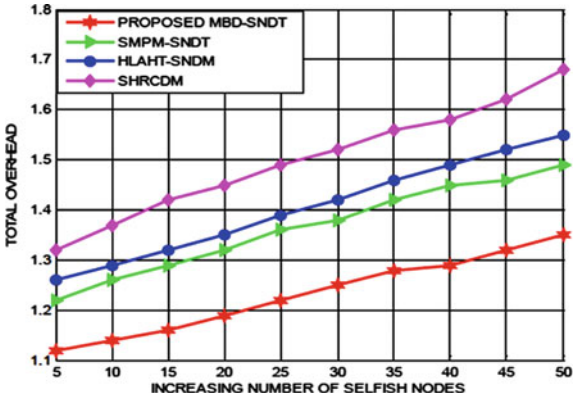



**Fig. 15** Performance of MBD-SNDT utilizing packet delivery ratio under expanding mobile nodes

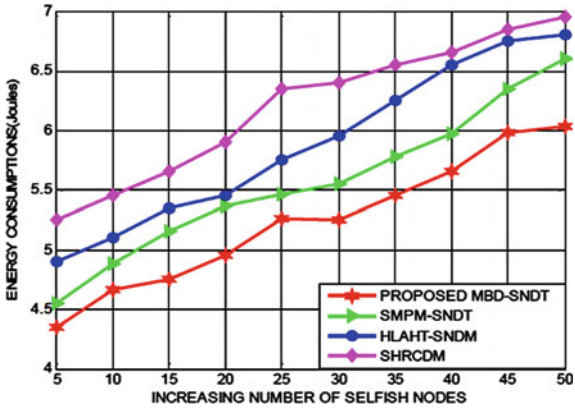



**Fig. 16** Performance of MBD-SNDT utilizing energy utilization underneath expanding mobile nodes

## 9 Conclusion

The MSD-SNDT was presented as a strong undertaking for basic acknowledgment of extremist direct by researching different degrees of incredible components that contribute toward convincing egotism area. This Proposed APPROACH was also assessed to be unmatched in the productive ID of SELFISH NODES through the multi-dimensional assessment. Each PORTABLE NODES and its characteristic toward the sending ability of other cooperating MOBILE NODES are monitored. The reenactment tests and delayed consequences of the PROPOSED MSD-SNDT approach were set out to be astounding in diminishing the control overhead, outright overhead, and PACKET torpidity on ordinary by 19, 15, and 17% overwhelming than the biased hub area plans used for analysis. The area movement of the PROPOSED MSD-SNDT approach was, moreover, asserted to be redesigned by 12% which was excellent to the saw biased NODE acknowledgment plans.

This proposed MBD-SNDT became the use of variance and mean packet forwarding rate for estimating the degree of effect. Through the selfish characteristics

of the mobile node in the direction of the network such that correct detection and isolation of selfish nodes can be imposed for reinforcing the degree of network overall performance. The simulation experiments of the planned MBD-SNDT revealed a median development rate of 19% and 17% in packet shipping and throughput with 20 and 12%. This minimizes the energy consumptions and routing overhead rate as compared to the selfish nodes isolation tactics contributed in the literature. The mean rate in the discovery of the proposed became improved via 18% superior on par with the compared selfish nodes isolation strategies.

## References

1. Santhosh Kumari D, Thirunadana Sikamani K (2015) Revival of selfish nodes in clustered MANET. Int J Adv Eng Technol 8(3):412-419. (ISSN: 2231-1963)
2. Fogue M, Garrido P, Martinez FJ, Cano JC, Calafate CT, Manzoni P (2015) CoCoWa: a collaborative contact-based watchdog for detecting selfish nodes. IEEE Trans Mob Comput 14(6):1162–1175. (P-ISSN: 1536-1233, E-ISSN: 1558-0660)
3. Xia H, Jia Z, Li X, Ju L, Sha EH-M (2013) Trust prediction and trust-based source routing in mobile ad hoc networks. Elsevier- Ad Hoc Netw 11, 2013, pp 2096–2114. (ISSN: 1570-8705)
4. Annamalai Giri A, Mohan E (2018) Distributed attack detection for wireless sensor networks. Int J Eng Technol 7(6):465–468. (ISSN: 2227-524X)
5. Zouridaki C, Mark B L, Hejmo M, Thomas RK (2005) A quantitative trust establishment framework for reliable data packet delivery in MANETs. In: Proceedings of ACM SASN, pp 1–10
6. Sengathir J, Manoharan R (2015) A futuristic trust coefficient-based semi- Markov prediction model for mitigating selfish nodes in MANETs. Springer Open J EURASIP J Wirel Commun Netw 2015(1):1–13. (ISSN: 1687-1499, 1687-1472)
7. Tarannum R, Pandey Y (2012) Detection and deletion of selfish MANET nodes-a distributed approach. In: 2012 1st International conference on recent advances in information technology (RAIT), IEEE Xplore May 2012, pp 45–56. (E-ISBN: 978-1-4577-0697-4:P-ISBN: 978-1-4577-0694-3)
8. Sengathir J, Manoharan R (2016) Exponential reliability factor based mitigation mechanism for selfish nodes in MANETs. J Eng Res 4(1):67–78. (ISSN: 2307-1877 Online ISSN: 2307-1885)
9. Karthikayen A, Selvakumar Raja S (2018) A skellam distribution inspired trust factor-based selfish node detection technique in MANETs. J Adv Res Dyn Control Syst JARDCS 10(13-Special Issue):940–949. ISSN: 1943-023X
10. Tamilarasan S, Aramudan M (2011) A performance and analysis of misbehaving node in MANET using intrusion detection system. IJCSNS Int J Comput Sci Netw Secur 11(5):258–264. ISSN: 1738-7906
11. Ali R, Griggio A, Franzén A, Dalpiaz F, Giorgini P (2012) Optimizing monitoring requirements in self-adaptive systems. In: International conference on exploring modeling methods for systems analysis and design, pp 362–377, book series—Lecture notes in business information processing book series LNBIP, vol 113. https://link.springer.com/chapter/10.1007/978-3-642-31072-0_25
12. Capkun S, Buttyan L, Hubaux JP (2003) Self-organized public-key management for mobile ad hoc networks. IEEE Trans Mob Comput 2(1):52–64. https://doi.org/10.1109/tmc.2003.1195151. P-ISSN: 1536-1233, E-ISSN: 1558-0660
13. Zhang Y (2011) An energy efficient-based AODVM routing in MANET. In: International conference on information and management engineering innovative computing and information, pp 66–72, book series—Communications in computer and information science CCIS, vol me 232

14. El-Bendary AM, Mohsen, Shen, Xuemin Sherman (2014) Secure routing protocols. In: Security for multi-hop wireless networks, pp 63–93. https://www.springer.com/gp/book/9783319046020
15. Cano J-C, Manzoni P, Kim D, Toh C-K (2007) A low-complexity routing algorithm with power control for self-organizing short-range wireless networks. Springer Wirel Pers Commun Int J 41:407–425. https://doi.org/10.1007/s11277-006-9150-6. EISSN 1572-834X, PISSN 0929-6212
16. Ramya K, Kavitha T (2016) Deterring selfish nodes using hierarchical account-aided reputation system in MANET. In: International conference on computing technologies and intelligent data engineering (ICCTIDE'16), pp 34–45. EISBN: 978-1-4673-8437-7, PISBN: 978-1-4673-8438-4. https://doi.org/10.1109/icctide.2016.7725351

# Metaheuristic-Enabled Shortest Path Selection for IoT-Based Wireless Sensor Network


**Subramonian Krishna Sarma**

**Abstract** IoT is defined as a pervasive and global network that aids and provides the system for monitoring and controlling the physical world through the processing and analysis of generated data by IoT sensor devices. Wireless sensor networks (WSNs) are comprised of a large number of nodes distributed in a vast region. Routing protocols are responsible for the development and the management of network routes. This paper intends to propose an optimized routing model for selecting the optimal shortest path in IoT-based WSN. More particularly, a dragonfly algorithm with Brownian motion (DABR) model is introduced to select the optimal route by taking into consideration of certain constraints such as (i) delay (ii) distance (iii) packet drop rate (PDR) and (iv) energy. Finally, the performance of the proposed work is compared with the conventional models to demonstrate the superior performance.

**Keywords** Wireless sensor network · Routing protocols · Dragonfly algorithm · Optimization · Brownian motion · Packet drop rate

## 1 Introduction

In day-to-day life, the IoT for smart cities has introduced vast technical improvements gradually that brings more comfortable and easiest life style [1]. Sensing and disseminating the information to the base station in a timely manner is the main feature of IoT-enabled applications [2]. The improvements in IoT aid in improving the communities by enhancing the infrastructure, providing more reliable and cost-effective municipal facilities, enhancing public transport, reducing road congestion and ensuring that citizens are healthy and more active in the society [3–5]. In an IoT model, the WSN is an essential feature. WSN is a network made up huge sensor nodes, in which every node has a sensor for detecting physical phenomena [6]. Here, the key challenge is to solve the problem of delivering sensed data with energy-efficient communication between sensor nodes through the shortest path routing [7,

S. K. Sarma (✉)
Working for UST, Europe, UK
e-mail: subramonian.ks@ieee.org



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_8

8]. A routing protocol seems to be a procedure for determining the appropriate path for the data transmission from source to destination. While selecting the route, this process faces several difficulties that depend on the network performance metrics, channel characteristics and types of networks [9]. Usually, the sensor node information in WSN will be sent to the base station linking the sensor network to some other networks, in which the data is collected, analysed and some action being taken if necessary [10]. The sensor nodes does not generate and distribute the information in multihop communication, but often acts as a route to the base station for many other sensor nodes [11]. The optimization algorithms find rapid usage in many engineering problems [12–14]. The major contribution of the research work is to propose the DABR algorithm for selecting the optimal shortest route in IoT for data sharing among nodes. In addition, the constraints like PDR, energy, distance and delay are considered for optimal route selection.

This paper deals with the following sections: Sect. 2 describes the reviews on optimal shortest path routing protocol in WSN. Section 3 deploys the proposed routing strategies in IoT. Section 4 provides the description on varied constraints for shortest route selection. Section 5 presents the optimal shortest path selection methodology via dragonfly algorithm with Brownian motion. Moreover, Sect. 6 portrays the results and their discussions, and in Section 7, conclusion of the research work is presented.

## 2 Literature Review

### *2.1 Related Works*

In 2018, Manu et al. [15] have projected a suitable way for routing based on IoT applications in WSN. Finally, the experimental outcomes of the proposed method indicated better end-to-end (e2e) delay, network throughput, energy level and packet delivery ratio. In 2020, Ahmad et al. [16] have introduced an energy-efficient geographic (EEG) routing protocol based on the performance and energy alert of routing in IoT permitted WSN. At last, the simulation results of the proposed method have shown better energy consumption and PDR than other methods. In 2019, Thangaramya et al. [17] have suggested a neuro-fuzzy and energy-aware cluster-founded routing algorithm in IoT for WSN. Finally, the experimental outcomes of the proposed method indicated superior performance in delay and network lifetime, packet delivery ratio and energy utilization. In 2019, Yu et al. [18] have determined a sector-based random routing (SRR) protection system in IoT for the privacy of the source site. At last, experimental outcomes of the proposed protocol have revealed its superiority with improved network lifetime and better security. In 2018, Guangjie et al. [19] have developed a source position security protocol for IoT using dynamic routing in WSNs. Finally, the simulation results of the proposed method have shown better performance than other traditional methods. In 2020, Tang et al. [20] have

presented the adaptive dual-mode routing-based mobile data gathering algorithm (ADRMDGA) in RWSNs for IoT. Finally, the experimental outcome shows that the ADRMDGA had better energy equilibrium as well as efficiently expanded the lifetime of the network. In 2018, Tang et al. [21] have presented the mathematical model designed for the innovative generation of promoting QoS routing determination. Finally, the simulation results reveal the presented method had more green efficiency over multihop IoT. In 2020, Deebak and Al-Turjman [22] have suggested the authentication and encryption model. Finally, the result of the presented model demonstrates that it had a superior percentage of monitoring nodes when compared with the conventional routing models. Table 1 depicts the features and challenges of IoT-based wireless sensor network using various techniques.

**Table 1** Features and challenges of IoT-based wireless sensor network using various techniques

| Author [citation] | Methodology | Features | Challenges |
|---|---|---|---|
| In 2018, Manu et al. [15] | Congestion and interference aware energy-efficient routing technique | Higher delivery rates | More improvement is needed with the optimization algorithm |
| In 2020, Ahmad et al. [16] | Mean square error algorithm | Reduces the energy holes in the network | Computational delay of the proposed work is higher |
| In 2019, Thangaramya et al. [17] | Convolution neural network with fuzzy rules | Higher network lifetime | Routing overhead |
| In 2019, Yu et al. [18] | The sector-based random routing scheme | Efficient protection for source location privacy | Privacy issue in multiple source nodes |
| In 2018, Guangjie et al. [19] | Source Location Protection Protocol Based On Dynamic Routing (SLPDR) | Improved network lifetime | Need a more effective protocol to protect the source location |
| In 2020, Tang et al. [20] | Dual-mode routing-based mobile data gathering algorithm | Improved energy equilibrium | The calculation amount is too large |
| In 2018, Hasan et al. [21] | Traffic system model designed with Markov discrete-time M/M/1 queuing model | High throughput | The lifetime of the network is low |
| In 2020, Deebak and Al-Turjman [22] | OLSR and AOMDV protocols | Less energy consumption and successful delivery factor | Not suitable for a high dynamic network |

# 3 Proposed Routing Strategies in IoT

## *3.1 Network Model*

The Internet of Things helps to connect the objects such as computing devices, machine and automation systems by using the internet. Figure 1 shows the principle of IoT concept, and it helps to identify and manage the things from anywhere in the world via Internet.

WSN is used to monitor and control various domestic and industrial automations. Emerging IoT devices makes it possible to develop cost-effective wireless sensor nodes with Internet connectivity. The combination of IoT and WSN is moving towards edge technology. In the conventional IoT-based WSN system, the radio frequency (RF) is used to send their information to the Internet. But, in the proposed network model, the WSN is connected to each IoT device, and the data is transmitted using the Internet protocol. The pictorial representation of IoT scenario is shown in Fig. 2.

Let $S_1, S_2, \ldots, S_n$ be the sensor nodes, in which $S_n$ represents the total number of nodes. Here, this scenario includes decentralized data centre (DDC) $P_1, P_2, \ldots, P_m$ and centralized data centre (CDC) $Q_1, Q_2, \ldots, Q_m$, where $P_m$ and $Q_m$ refer to the total number of connected DDCs and CDCs, respectively. Throughout data transmission process, the receiver is decided on the basis of two perspectives: if DDC is the destination, the receiver who received the information from the source is considered to be its neighbourhood node. If the endpoint is CDC, it serves as a receiver; however, the node does not specify the PDR as well as energy values. The presented


**Fig. 1** Components of IoT

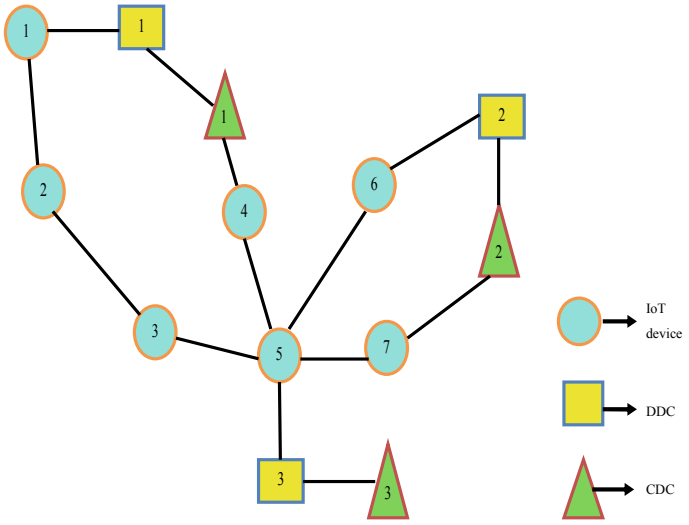



**Fig. 2** Pictorial representation of IoT scenario (IoT device is nothing but sensor)

work focuses on optimal shortest route selection among all the nodes with safe transmission of data. The nodes are configured with the specifications including distance ($D$), delay (Del), energy ($E$) and PDR for data transmission. Here, a new single-objective function would be depending on these limitations to choose the optimal route.

## 4 Description on Varied Constraints for Shortest Route Selection

### *4.1 Distance Model*

Consider two sensor nodes such as $s$(normal node) and $r$(another node) and their locations be $u$ and $v$, respectively. The distance $D$ among the consecutive nodes can be represented by Eq. (1), which portrays the Euclidean distance among the nodes.

$$D = \sqrt{(r_u - s_u)^2 + (r_v - s_v)^2} \tag{1}$$

### 4.2 Delay

The delay can be defined as the ratio of distance divided by the data speed $H$ transmission. The formulation for delay Del can be expressed using Eq. (2) [23].

$$\text{Del} = \frac{D}{H} \tag{2}$$

### 4.3 Energy Model

Energy utilization remains a big problem when transmitting the data in IoT. In addition, the consumed battery could not be re-energized; thereby, if the battery is low, the transmission could fail. Usually, the network absorbs extra energy by performing a different function such as sensing, aggregating, transmitting and receiving. Here, the design of energy requirement used in data transmission is given in Eq. (3), where $E_g$ represents the electronic energy depending on diverse constraints, $E_{fs}$ signifies the energy essential while using free space, $E_{pw}$ indicates energy of the power amplifier and $E_T(M : f)$ determines the total utilized energy required for transferring $M$ bytes of packets at distance $f$. The model for electronic energy can be represented using Eq. (4) where c denotes the utilized energy during data aggregation. The total utilized energy required for attaining $M$ bytes of packets at distance $f$ can be determined in Eqs. (5), and (6) indicates the energy required for amplification $E_a$.

$$E_T(M : f) = \begin{cases} E_g * M + E_{fs} * M * f^2, if \ f < f_0 \\ E_g * M + E_{pw} * M * f^2, if \ f \geq f_0 \end{cases} \tag{3}$$

$$E_g = E_T + E_d \tag{4}$$

$$E_R(M : f) = E_g M \tag{5}$$

$$E_a = E_f f^2 \tag{6}$$

$$f_0 = \sqrt{\frac{E_f}{E_p}} \tag{7}$$

In Eq. (3), the threshold distance $f_0$ is evaluated as per Eq. (7), in which $E_p$ indicates the energy of power amplifier and $E_f$ determines the needed energy while exploiting the free space model. On the whole, the entire network energy can be represented by Eq. (8), and hence, $E_{\text{id}}$ indicates the energy required during the idle

state, $E_R$ determines the total energy received and $E_{\text{sen}}$ denotes the cost of energy during the sensing process. However, it is essential to decrease the total energy as exposed in Eq. (8).

$$E_{\text{tot}} = E_T + E_R + E_{\text{id}} + E_{\text{sen}} \tag{8}$$

### 4.4 PDR

**SF attack and black hole attack**: Here, the malevolent sensor node transfers only specific data packets in the SF attack and removes some other packets. The malicious node declines all packets earned without transferring them in the event of a black hole attack [24]. If the PDR on a node $C_i$ goes beyond the threshold value $\chi_{pdr_{\text{SF}}}$, here $C_i$, the thought can be taking those SF attack. In addition, if the PDR value of node $C_i$ goes further than the threshold values $\chi_{pdr_{\text{BH}}}$, respectively, thus it forms a blackhole attack.

**Wormhole attack**: Throughout this wormhole attack, the malicious node moves all the data packets through the tunnel to another malicious node, rather than transferring to another legal node [25]. If PDR rate of $C_i's$ nearest node is better than the threshold cost $\chi_{pdr_{\text{WH}}}$ of the node $C_i$, then $C_i$ takes out a wormhole attack.

## 5 Optimal Shortest Path Selection via Dragonfly Algorithm with Brownian Motion

### 5.1 Objective Function

The proposed research work on shortest path routing focuses on identifying the optimal path besides data communication, taking into account the significant challenges including such delay, distance, PDR and energy. The objective function of the proposed work seeks to reduce the distance between consecutive nodes and also at reducing the delay in transferring data from one node to the next. Furthermore, the security threats and packet drop rates should be minimized for better system performance. On the other hand, the network energy should be greater indicating that this would only use a limited amount of energy when transferring the data.

Here, the defined single-objective function is expressed in Eq. (12), where the fitness functions such as $F_1$,$F_2$ and $F_3$ are calculated as per Eqs. (9), (10) and (11). However, the values of $\alpha$, $\beta$ and $\delta$ are fixed as 0.8.

$$F_1 = (\alpha \times D) + (1 - \alpha) \times \text{Del} \tag{9}$$

$$F_2 = (\beta \times F_1) + (1 - \beta) \times \text{PDR} \tag{10}$$

$$F_3 = \delta \times F_2 + (1 - \delta) \times \frac{1}{E} \tag{11}$$

$$F = \text{Min}(F_3) \tag{12}$$

### 5.2 Solution Encoding

The solution specified in the adopted method is the nodes as demonstrated in Fig. 3. Here,$K_N$ indicates the sum number of nodes ($K_N = 1000$). The maximum and minimum bounds are provided as $X_{\max}$ and $X_{\min}$ that may be either 1's or 0's. Consequently, the nodes can be represented as 1's for selected data transmission, whereas the remaining nodes that fall with 0's are dropped.

### 5.3 Proposed DABR Algorithm

However, the existing Dragonfly algorithm (DA) model results in effective approximations; this also includes few disadvantages like low internal memory and slow convergence [26]. The Brownian motion seems to be the random motion of suspended liquid molecules arising through collisions of quickly flowing fluid molecules [27]. Therefore, to prevail over the drawbacks of existing DA, some improvements are made in the proposed work by integrating the concept of Brownian motion with the proposed work. The proposed DA process involves the following steps with two significant stages: exploitation and exploration.

The variation of DA using the Brownian motion can be determined in Eqs. (13), (14) and (15).

$$Y_{t+1} = Y_t + h * \text{rand}() \tag{13}$$

$$h = \sqrt{\frac{TP}{N}} \tag{14}$$

$$N = 100 * TP \tag{15}$$

In Eq. (14), the $TP$ indicates the motion time period in seconds of an agent in dragonfly. Moreover, the $TP$ rate can be chosen as 0.01 and $N$ denotes the number of sudden motions in proportion to time for a similar agent. Figure 4 shows the flowchart

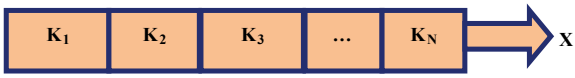



**Fig. 3** Solution encoding

**Table 2** Simulation parameters

| Methods | Parameters | Value |
|---|---|---|
| Brownian motion | Motion time | 0.01 s |
| | $n$ | $n$ = 100*motion time |

representation of the proposed DABR algorithm. In contrast to conventional methods like particle swarm optimization (PSO), the proposed DABR updates the position by both attraction and diffraction, whereas the standard PSO considers only attraction-based update.

## 6 Results and Discussion

### *6.1 Simulation Procedure*

The proposed method for optimal shortest path selection using DABR was implemented in MATLAB, and the following results were achieved. The proposed DABR method was evaluated over other traditional methods like PSO [28], grey wolf optimization (GWO) [29] and DA [26], and the their outcomes were observed in terms of **cost function, delay, distance, energy and PDR**. Here, 1000 nodes were considered, and among that 78th node was taken as source node and 900th node was taken as destination node. Here for analysis purpose, the location, PDR and energy were varied for two sets of values that were considered as Data 1 and Data 2 in results section. Moreover, the analysis was carried out by varying the count of iterations from 0, 20, 40, 60, 80 and 100 (Table 2).

### *6.2 Performance Analysis*

The performance of the proposed DABR model for optimal shortest path routing selection in terms of cost function, delay, distance, energy and PDR for Data 1 and Data2 is given in Figs. 5 and 6. Here, the analysis was carried out based on various iterations such as 0, 20, 40, 60, 80 and 100, respectively. From the graphical analysis, the proposed DABR method obtains lower values for cost function, delay, distance, energy and PDR with increased in iteration. In Fig. 5a, the cost function of the proposed DABR method at 100th iteration is 6.77%, 5.08% and 4.23% better than the traditional methods such as PSO, GWO and DA, respectively. Further in Fig. 5b,

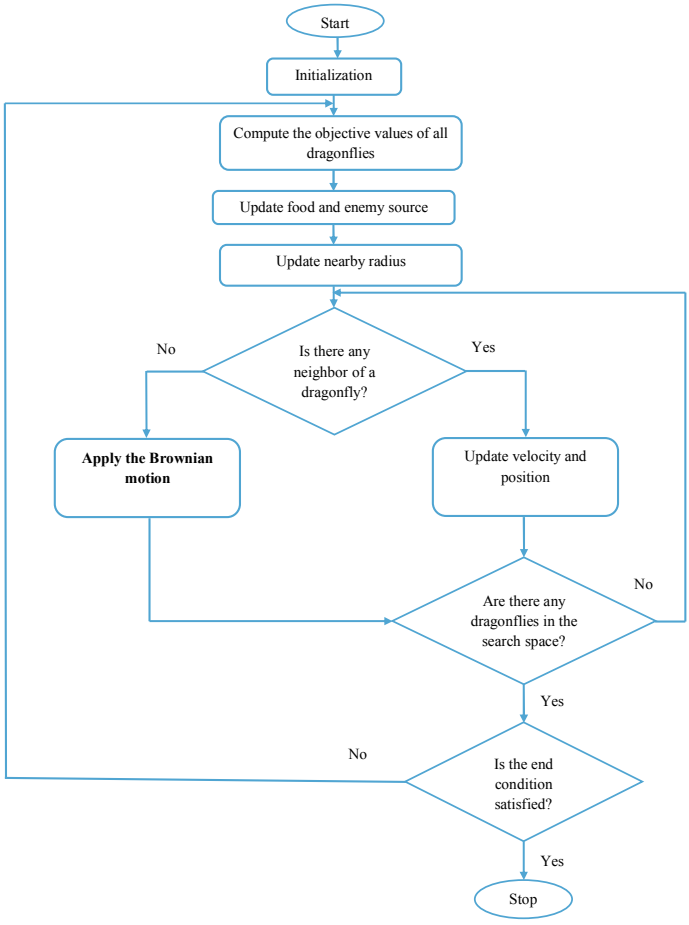



**Fig. 4** Flowchart of the DABR algorithm

the delay of the proposed DABR method at 80th iteration is 7.69, 5.64 and 7.69% better than the existing models like PSO, GWO and DA. In Fig. 5c, at 100th iteration, the distance measure of the proposed DABR method obtains the value of 0.35 which is superior to the existing models such as PSO, GWO and DA that holds the values of 0.38, 0.37 and 0.384, respectively. However, in Fig. 5d, the energy measure of the proposed DABR method at 60th iteration is 6.80, 5.23 and 2.35% better than PSO, GWO and DA models. Also, in Fig. 5e, the PDR values of the proposed DABR method at 80th iteration is 0.12 that is better than the existing models like PSO and GWO with the values of 0.128 and 0.133. From the graphical analysis on Data 2, the proposed DABR method at 100th iteration obtains better values for cost function, delay, distance, energy and PDR. In Fig. 6a, the cost function of the proposed DABR method at 100th iteration is 10.85%, 11.38% and 11.90% superior to the traditional

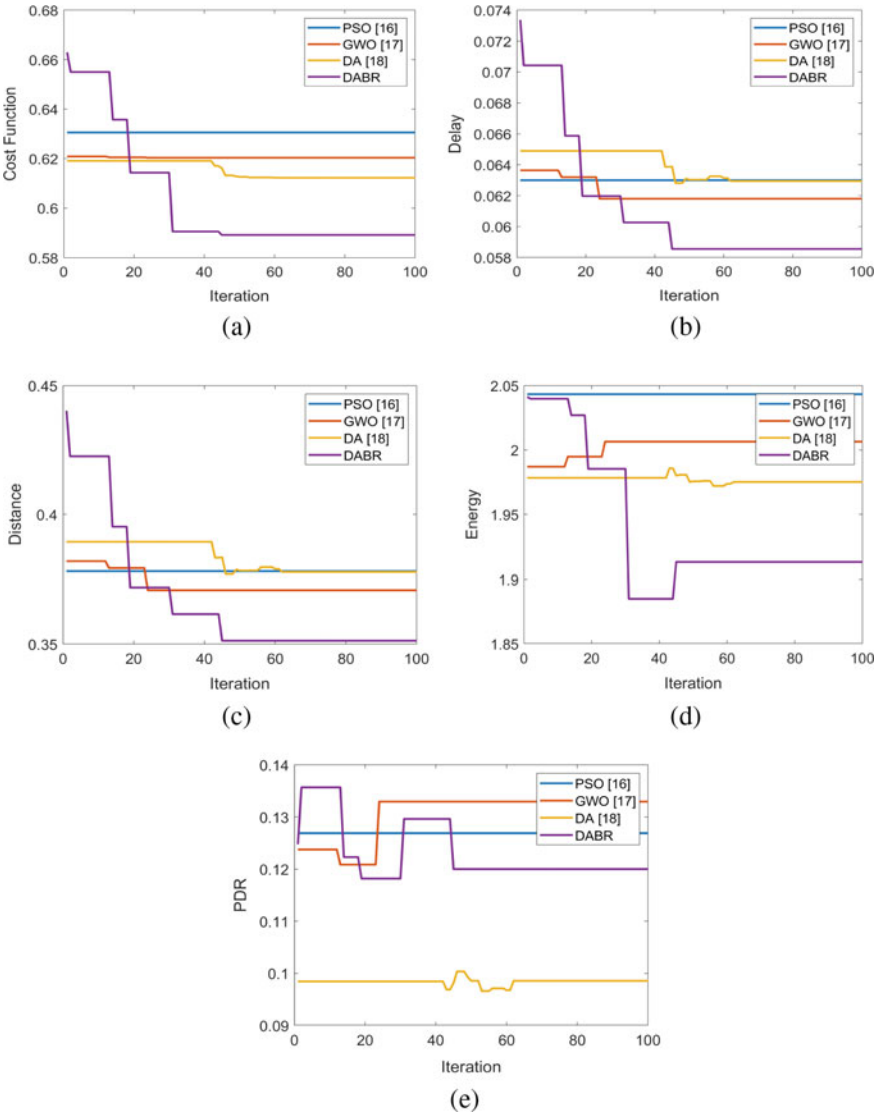



**Fig. 5** Performance analysis of the proposed method over the traditional models with respect to constraints such as **a** cost function, **b** delay, **c** distance, **d** energy, **e** PDR for data 1

models such as PSO, GWO and DA, respectively. In Fig. 6b, the delay value of the proposed DABR method obtains 0.0626, whereas the compared existing models like PSO, GWO and DA hold the values of 0.064, 0.0674 and 0.065 at 80th iteration. The distance measure of the proposed DABR method is 0.7%, 5.4% and 2.08% better than the existing models such as PSO, GWO and DA models at 100th iteration as shown in Fig. 6c. Further, in Fig. 6d, the energy utilization of the proposed DABR

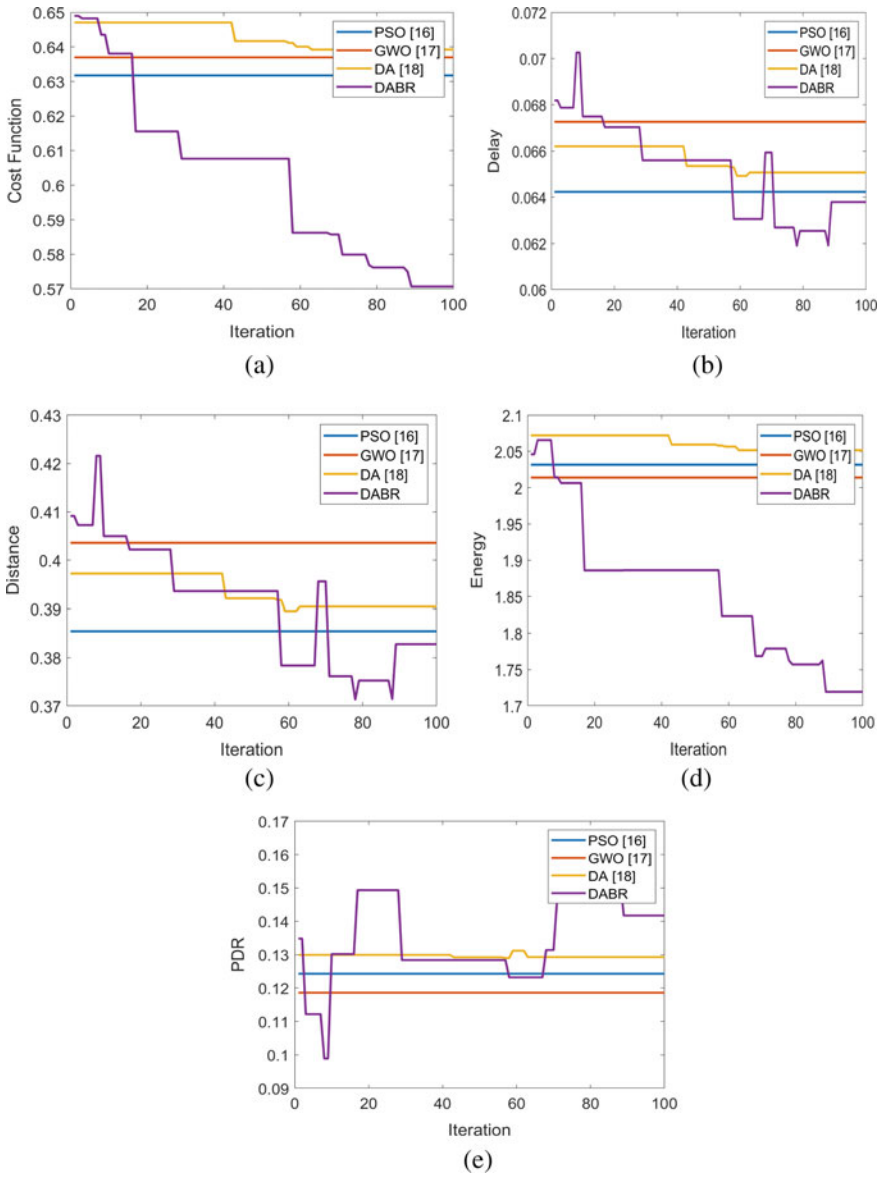



**Fig. 6** Performance analysis of the proposed method over the traditional models with respect to constraints such as **a** cost function, **b** delay, **c** distance, **d** energy, **e** PDR for data 2

**Table 3** Statistical analysis of the proposed DABR method over other traditional models

| Methods | PSO [28] | GWO [29] | DA [26] | DABR |
|---|---|---|---|---|
| Best | 0.63056 | 0.62033 | 0.61224 | 0.5707 |
| Worst | 0.63175 | 0.637 | 0.63923 | 0.58918 |
| Mean | 0.63115 | 0.62866 | 0.62574 | 0.57994 |
| Median | 0.63115 | 0.62866 | 0.62574 | 0.57994 |
| STD | 0.00084656 | 0.011787 | 0.019085 | 0.013071 |

method is 18.31, 17.44 and 19.18% better than PSO, GWO and DA models at 100th iteration. Also, in Fig. 6e, the PDR value of the proposed DABR method at 80th iteration is 0.148, which is lower than the values attained by existing models such as PSO, GWO and DA. Thus, the improvement of the presented DABR model is proved over other traditional models.

## *6.3 Statistical Analysis*

The statistical analysis of the proposed DABR method over other traditional models is given in Table 3. Since metaheuristic algorithms are stochastic in nature, the simulation is carried out for two times and the results are taken. The proposed DABR method obtains better outcomes over the existing methods such as PSO, GWO and DA. For bestcase scenario, the proposed DABR method is 10.48%, 8.69%, and 7.27% better than the existing models like PSO, GWO and DA, correspondingly. Further, the proposed DABR method for worst-case scenarios 7.22%, 8.11% and 8.49% is lower than the traditional models such as PSO, GWO and DA, correspondingly. The proposed DABR method for both mean and median case scenarios is 8.83, 8.40 and 7.89% lesser than the existing models like PSO, GWO and DA. Also, the proposed DABR method for standard deviation scenario is 0.013071 that is lower than the values obtained by PSO(0.00084656) and GWO (0.011787), respectively. Thus, the proposed DABR methods have shown better outcomes when compared to other traditional models.

## *6.4 Overall Analysis of the Proposed Model*

Tables 4 and 5 determine the overall analysis of the proposed DABR method over the traditional models for both Data 1 and Data 2, respectively. Here, the proposed DABR method obtains lower values than the other existing models in terms of total cost, distance, delay, PDR and energy. Table 4 demonstrates the overall performance analysis of the proposed DABR method over the existing models for Data 1. The total cost for the proposed DABR method is 7.02%, 5.28% and 3.91% better than the

**Table 4** Overall analysis of the proposed DABR method over other traditional models for data 1

| Metrics | PSO [28] | GWO [29] | DA [26] | DABR |
|---|---|---|---|---|
| Total cost | 0.63056 | 0.62033 | 0.61224 | 0.58918 |
| Distance | 0.37798 | 0.3708 | 0.37768 | 0.35132 |
| Delay | 0.062996 | 0.0618 | 0.062947 | 0.058553 |
| PDR | 0.12689 | 0.13295 | 0.098566 | 0.12 |
| Energy | 2.0433 | 2.0065 | 1.9752 | 1.9131 |

**Table 5** Overall analysis of the proposed DABR method over other traditional models for data 2

| Metrics | PSO [28] | GWO [29] | DA [26] | DABR |
|---|---|---|---|---|
| Total cost | 0.63175 | 0.637 | 0.63923 | 0.5707 |
| Distance | 0.3854 | 0.40358 | 0.39042 | 0.38275 |
| Delay | 0.064234 | 0.067264 | 0.06507 | 0.063792 |
| PDR | 0.12427 | 0.11858 | 0.12928 | 0.14173 |
| Energy | 2.0316 | 2.0139 | 2.0516 | 1.7194 |

traditional methods such as PSO, GWO and DA, respectively. Further, both distance and delay of the proposed DABR method are 7.58, 5.54 and 7.50% better than the existing models like PSO, GWO and DA. For PDR value, the proposed DABR method is 0.12 superior to the existing models such a PSO (0.12689) and GWO (0.13295). Moreover, the energy utilized for the proposed DABR method is 6.80%, 4.88% and 3.246% better than existing models like PSO, GWO and DA, respectively.

Similarly from Table 5, the total cost of the proposed DABR method is 10.69, 11.61 and 12% better than the traditional methods such as PSO, GWO and DA models for Data 2. Moreover, both distance and delay of the proposed DABR method is 0.6, 5.44 and 2% superior to the existing models like PSO, GWO and DA. The PDR value of the proposed DABR method is found to be 0.14173, which is better than the existing models such as PSO, GWO and DA that holds comparatively higher values of 0.12427, 0.11858 and 0.12928, correspondingly. Finally, the energy utilized by the proposed DABR method is 18.15%, 17.12% and 19.32% superior to the existing models like PSO, GWO and DA, respectively. Thus, the proposed DABR method was found to offer better performance including the total cost, distance, delay, PDR and energy.

## 7 Conclusion

An optimized routing model for selecting the optimal shortest path in IoT-based WSN is presented in this research work using DABR model. The optimal route is obtained by considering the parameters such as PDR, distance, delay and energy.

From the analysis, the cost function of the proposed DABR method is observed as 6.77, 5.08 and 4.23% which is better than the traditional methods such as PSO, GWO and DA. In addition, the PDR value of the proposed DABR method is obtained as 0.14173, which is much better than the existing models such as PSO, GWO and DA that holds comparatively higher values of 0.12427, 0.11858 and 0.12928, correspondingly. Finally, the energy utilized for the proposed DABR method is 18.15%, 17.12% and 19.32% better than existing models like PSO, GWO and DA, respectively. In the IoT-based WSN, if the practical constraints of security and QoS measures such as throughput and packet delivery ratio are considered, the objective model becomes more complicated. Hence, in this work, physical constraints such as PDR, energy distance and delay are considered for optimal route selection. In addition, attack detection is also framed and it can be framed in two ways such as rule-based approach [30] and predictive approach [31]. In future, this work will be extended to solve the above-mentioned issues in IoT for data sharing among nodes.

## References

1. Singh R, Verma AK (2017) Energy efficient cross layer based adaptive threshold routing protocol for WSN. AEU I J Electr Commun 72:166–173
2. Ke W, Yangrui O, Hong J, Heli Z, Xi L (2016) Energy aware hierarchical cluster-based routing protocol for WSNs. J China U Posts Telecommun 23(4):46–52
3. Hong C, Zhang Y, Xiong Z, Xu A, Ding W (2018) FADS: circular/spherical sector based forwarding area division and adaptive forwarding area selection routing protocol in WSNs. Ad Hoc Network 70:121–134
4. Mujica G, Portilla J, Riesgo T (2015) Performance evaluation of an AODV-based routing protocol implementation by using a novel in-field WSN diagnosis tool. Microprocess Microsyst 39(8):920–938
5. Misra G, Kumar V, Agarwal A, Agarwal K (2016) Internet of things (iot)–a technological analysis and survey on vision, concepts, challenges, innovation directions, technologies, and applications (an upcoming or future generation computer communication system technology). Am J Electr Electron Eng 4(1):23–32
6. Bhardwaj R, Kumar D (2019) MOFPL: multi-objective fractional particle lion algorithm for the energy aware routing in the WSN. Pervasive Mob Comput 58:
7. Rani S, Malhotra J, Talwar R (2015) Energy efficient chain based cooperative routing protocol for WSN. Appl Soft Comput 35:386–397
8. Behera TM, Mohapatra SK, Samal UC, Khan MS (2019) Hybrid heterogeneous routing scheme for improved network performance in WSNs for animal tracking. Internet Things 6:
9. Yarinezhad R, Hashemi SN (2019) Solving the load balanced clustering and routing problems in WSNs with an fpt-approximation algorithm and a grid structure. Pervasive Mob Comput 58:
10. Fu X, Fortino G, Pace P, Aloi G, Li W (2020) Environment-fusion multipath routing protocol for wireless sensor networks. Inform Fusion 53:4–19
11. Toor AS, Jain AK (2019) Energy aware cluster based multi-hop energy efficient routing protocol using multiple mobile nodes (MEACBM) in wireless sensor networks. AEU I J Electr Commun 102:41–53
12. Singh G, Jain VK, Singh A (2018) Adaptive network architecture and firefly algorithm for biogas heating model aided by photovoltaic thermal greenhouse system. Energ Environ 29(7):1073–1097

13. Preetha NSN, Brammya G, Ramya R, Praveena S, Binu D, Rajakumar BR (2018) Grey wolf optimisation-based feature selection and classification for facial emotion recognition. IET Biometrics 7(5):490–499. https://doi.org/10.1049/iet-bmt.2017.0160
14. Jadhav AN, Gomathi N (2019) DIGWO: hybridization of dragonfly algorithm with Improvedc grey wolf optimization algorithm for data clustering. Multimedia Res 2(3):1–11
15. Elappila M, Chinara S, Parhi DR (2018) Survivable path routing in WSN for IoT applications. Pervasive Mob Comput 43:49–63
16. Hameed AR, Islam S, Raza M, Khattak HA (2020) Towards energy and performance aware geographic routing for IoT enabled sensor networks. Comput Electr Eng 85:
17. Thangaramya K, Kulothungan K, Logambigai R, Selvi M, Kannan A (2019) Energy aware cluster and neuro-fuzzy based routing algorithm for wireless sensor networks in IoT. Comput Network 151:211–223
18. He Y, Han G, Wang H, Ansere JA, Zhang W (2019) A sector-based random routing scheme for protecting the source location privacy in WSNs for the Internet of Things. Future Gener Comput Syst 96:438–448
19. Han G, Zhou L, Wang H, Zhang W, Chan S (2018) A source location protection protocol based on dynamic routing in WSNs for the social internet of things. Future Gener Comput Syst 82:689–697
20. Tang L, Guo H, Wu R, Fan B (2020) Adaptive dual-mode routing-based mobile data gathering algorithm in rechargeable wireless sensor networks for internet of things. Appl Sci 10(5):1821
21. Hasan MZ, Al-Turjman F, Al-Rizzo H (2018) Analysis of cross-layer design of quality-of-service forward geographic wireless sensor network routing strategies in green internet of things. IEEE Access 6:20371–20389
22. Deebak BD, Al-Turjman F (2020) A hybrid secure routing and monitoring mechanism in IoT-based wireless sensor networks. Ad Hoc Netw 97:102022
23. Kumar R, Kumar D (2016) Hybrid swarm intelligence energy efficient clustered routing algorithm for wireless sensor networks. J Sens
24. Sedjelmaci H, Senouci SM, Feham M (2013) An efficient intrusion detection framework in cluster-based wireless sensor networks. Secur Commun Network 6(10):1211–1224
25. Abduvaliyev A, Lee S, Lee YK (2010) Energy efficient hybrid intrusion detection system for wireless sensor networks. In: International conference on electronics and information engineering, vol 2, pp 25–29
26. Mirjalili1 S (2015) Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput. Appl. 27(4):1053–1073
27. Acı ÇI, Gulcan H (2019) A modified dragonfly optimization algorithm for single- and multiobjective problems using Brownian motion. Comput Intell Neurosci 17: https://doi.org/10.1155/2019/6871298
28. Wang D, Tan D, Liu L (2017) Particle swarm optimization algorithm: an overview. Soft Comput 22(2):387–408
29. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
30. Li X, Yuan J, Ma H, Yao W (2018) Fast and parallel trust computing scheme based on big data analysis for collaboration cloud service. IEEE Trans Inform Forensics Secur 13(8):1917–1931
31. Krishna SS (2019) Optimized activation function on deep belief network for attack detection in IoT. In: 2019 Third international conference on I-SMAC (IoT in social, mobile, analytics and cloud) (I-SMAC), pp 702–708. IEEE

# Improved Harris Hawks Optimization Algorithm for Workflow Scheduling Challenge in Cloud–Edge Environment


**Miodrag Zivkovic, Timea Bezdan, Ivana Strumberger, Nebojsa Bacanin, and K. Venkatachalam**

**Abstract** Edge computing is a relatively novel technology, which is closely related to the concepts of the Internet of things and cloud computing. The main purpose of edge computing is to bring the resources as close as possible to the clients, to the very edge of the cloud. By doing so, it is possible to achieve smaller response times and lower network bandwidth utilization. Workflow scheduling in such an edge–cloud environment is considered to be an NP-hard problem, which has to be solved by a stochastic approach, especially in the scenario of multiple optimization goals. In the research presented in this paper, a modified Harris hawks optimization algorithm is proposed and adjusted to target cloud–edge workflow scheduling problem. Simulations are carried out with two main objectives—cost and makespan. The proposed experiments have used real workflow models and evaluated the proposed algorithm by comparing it to the other approaches available in the recent literature which were tested in the same simulation environment and experimental conditions. Based on the results from conducted experiments, the proposed improved Harris hawks optimization algorithm outperformed other state-of-the-art approaches by reducing cost and makespan performance metrics.

**Keywords** Cloud–edge computing · Harris Hawks optimization · Swarm intelligence · Workflow scheduling

M. Zivkovic (✉) · T. Bezdan · I. Strumberger · N. Bacanin
Singidunum University, Danijelova 32, 11000 Belgrade, Serbia
e-mail: mzivkovic@singidunum.ac.rs

T. Bezdan
e-mail: tbezdan@singidunum.ac.rs

I. Strumberger
e-mail: istrumberger@singidunum.ac.rs

N. Bacanin
e-mail: nbacanin@singidunum.ac.rs

K. Venkatachalam
School of Computer Science and Engineering, VIT Bhopal University, Bhopal, India
e-mail: venkatachalam.k@vitbhopal.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_9

## 1 Introduction

The Internet of things (IoT) is a novel technology that incorporates numerous remote devices connected via the Internet. These remote devices can collect data from their surroundings, process that data and share it with other devices in the network. The IoT includes several technologies, such as computing and communications, which should be carefully integrated in order to provide stable, reliable, and robust service to the users [1]. Some examples of such technologies are cloud computing, wireless sensor networks (WSN), radio frequency identification devices (RFID), and sensor technology.

Moreover, the IoT incorporates artificial intelligence methods and algorithms, that enable the connected devices to collaborate together and learn from previous experiences without human supervision. The IoT's nature is even more complex, as it also requires interaction with humans, and must meet users' demands while being able to dynamically adapt to the environment. Latency in data processing presents a considerable challenge, as in numerous applications data that has been collected and processed by the IoT devices is not valid after a defined time period. Latency is caused by the large amount of data processed by the IoT applications. Therefore, IoT must allow data collection and processing in real time. This challenge can be addressed by implementing edge computing in the cloud environment. The edge computing paradigm relies on the idea to bring the resources close to the end users, which in turn allows efficient handling of the tasks sensitive to the latency. On the other hand, cloud computing is flexible and scalable. Together, edge computing and cloud computing technologies complement each other [2]. The general architecture of the cloud–edge computing model consists of three layers [3]. Smart IoT devices (such as sensors and mobile phones) which are located in the end user layer send a request for services to the edge layer. Edge devices can process time-sensitive tasks and minimize latency. Edge layer is on the other side connected to the cloud system layer which can provide on-demand resources and data storage.

The biggest challenge in edge computing is considered finding an efficient task scheduling algorithm, which should minimize the task schedule duration and completion time. The problem of workflow scheduling in the edge computing is an NP-hard problem by nature, and an exhaustive and thorough search through the entire solution space cannot be performed. Unfortunately, as the amount of computing tasks rises, solution space grows exponentially, making it practically impossible to find the optimal solution in the acceptable time frame [4]. As the result, different stochastic approaches, such as metaheuristics, must be used to solve this problem.

This paper has devised an upgraded Harris hawks optimization (HHO) algorithm, that belongs to the swarm intelligence family, to specifically target the problem of workflow applications scheduling in hybrid cloud–edge environments, with a goal to keep the balance between makespan and cost.

The remainder of this paper is structured in the following way: Sect. 2 presents a survey of the recent metaheuristics and heuristics applications for workflow scheduling in cloud–edge surroundings, Sect. 3 provides insights of the utilized cloud–edge

model. Section 4 provides the details of the original and enhanced HHO approaches. Section 5 exhibits the simulation results and discussion, while Sect. 6 provides a conclusion and final remarks of this research along with the future work.

## 2 Background and Related Work

The main goal of any task scheduling algorithm is the minimization of performance metrics of the execution environment, such as the duration of all task execution, average response times, costs, and so on. Numerous heuristics that already exist can be used to improve task scheduling efficiency. Some of these approaches include first come first served (FCFS) heuristic, last come first served method (LCFS), join the shortest queue (JSQ) policy, heterogeneous earliest finish time approach (HEFT), etc.

However, since the workflow scheduling challenge is part of the NP-hard category, heuristics, as well as other classical deterministic approaches, cannot achieve satisfying results within a reasonable time frame. Therefore, the utilization of metaheuristics is a necessity. In general, metaheuristics are separated into those that are not inspired and the approaches that are motivated by nature (bio-inspired). Distinguished representatives of nature-inspired algorithms include evolutionary algorithm (EA) and swarm intelligence metaheuristics.

The most famous exemplar of the evolutionary algorithms is a genetic algorithm (GA). The GA has already been utilized with great success in solving numerous real-life NP-hard problems, including cloud computing load balancing and task scheduling [5, 6].

The second large group of nature-inspired algorithms, known as swarm intelligence metaheuristics, is inspired by social behavior expressed by the large groups of otherwise simple individuals such as ants, moths, fireflies, fish, wolves, and beetles. When these individuals are in the swarms, they can manifest very intelligent social behavior and coordinated movement, and this property inspired a huge variety of different swarm intelligence algorithms [7]. Swarm algorithms were successfully validated for many benchmark [8–10] and real-life challenges [11, 12].

Some of the swarm intelligence approaches that were successfully implemented and tested for cloud–edge scheduling problems in original and modified/hybridized versions include: bat algorithm (BA) [13], particle swarm optimization (PSO) [14], [2], cuckoo search (CS) [15], monarch butterfly optimization (MBO) [16], and whale optimization algorithm (WOA) [17].

## 3 Model Formulation

Edge–cloud computing is a novel approach that offers numerous new possibilities, but since it is a compound domain with a high amount of different types of resources, it still has some serious challenges which must be taken into the consideration. The

first challenge is the process of scheduling business tasks, which should be executed in real time, in the environment with a large amount of data flowing around. To address this challenge, it is necessary to choose a scheduling algorithm which will be efficient in making all business tasks being finished on time, which in turn will result in real-time execution.

The second major challenge for edge–cloud environment is handling business workflows, where workflows should be completed with respect to the QoS requirements (typically deadline and cost). Some tasks can be flexible in the scope of resource allocation, while other tasks require real-time execution on specific resources, and are considered to be not flexible. The selected scheduling algorithm should be capable to efficiently determine the appropriate resource allocation for each task, and at the same time to minimize the cost and task completion deadline.

Research conducted in [2] provided an excellent example which demonstrates the typical business workflow scheduling problem by considering a video surveillance/object tracking (VSOT) application on a use case of mobile and geo-distribution requirements. The proposed research work uses the same workflow scheduling problem formulation as this paper. The problem of workflow scheduling in the cloud–edge system, within the described context of VSOT application, can be formulated as follows—how to assign the resources (which can be of different types and with heterogeneous processing capabilities) to the individual tasks of the observed workflow with a goal to minimize the overall time required for the completion of all tasks and overall cost. Workflow is a sequence of tasks, which is usually observed in the shape of a directed acyclic graph (also known under the acronym DAG) and in turn, the DAG model can be used to directly express a typical workflow application.

Each task is denoted as one node in DAG, while the edges of the graph represent dependencies between the tasks. From a mathematical point of view, DAG exhibits the following properties: it is a finite graph with directed edges and without any directed cycles. Edges in DAG represent the sequences of execution. With these properties of DAG, it can be utilized to graphically denote various business activities together with the mutual dependencies that exist between the individual activities, namely the partial ordering and a sequence of execution.

The formal definition of the DAG can be summarized as follows: it is a graph $G = (T, E)$, where $T$ represents a collection of tasks of a workflow, and $E$ denotes the set of mutual time-related dependencies and possible communication constraints among the individual tasks. For every individual task in the collection of tasks, $t_s \in T(T = \{t_1, t_2, \ldots, t_n\})$ has assigned value of computational workload $cw_s$. A directed edge $e_{ij} = (t_i, t_j)$ means that the task $t_i$ must be finished before the start of the execution of the task $t_j$. Each edge $e_{ij}$ is assigned a weight, $cv_{ij}$, a non-negative value which marks the quantity of data transfer from the task $t_i$ to the task $t_j$. The workflow has one starting node, the task without direct predecessors, which is denoted with $t_{\text{start}}$. The workflow ends with task which does not have any direct successors, and it is denoted with $t_{\text{end}}$. Task $t_{\text{end}}$ is not possible to be executed unless all directly preceding tasks have been finished.

In the observed VSOT application scenario, makespan is defined as the maximum time for executing the workflow. Cloud–edge resources are split into two separate

groups, where the first group includes cloud servers, while the second group consists of edge servers. Every task of the workflow can be executed on either type of servers, and this is depending on the selected resource allocation approach. After allocating task $t_s$ to a specific server, the time required for the execution of the task $t_s$ can be calculated as given in Eq. (1):

$$T_{t_s}^{l} = \frac{cw_s}{\delta_l}, \tag{1}$$

where $\delta_l$ marks the processing power of the $l-th$ server. If it has been considered that $ct(e_{ij}^{l})$ denotes the time which is required for data transfer from task $t_i$ to the task $t_j$, it is possible to be calculated by using Eq. (2):

$$ct(e_{ij}^{l}) = \frac{cv_{ij}}{B} \tag{2}$$

Parameter $B$ in Eq. (2) marks the bandwidth of the links connecting the servers on which tasks have been allocated. Every individual task $t_s$ is allocated to only one server, it has the starting time $\mathrm{ST}_{t_s}$, specified with Eq. (3), and finishing time $\mathrm{FT}_{t_s}$, which can be calculated by using Eq. (4).

$$\mathrm{ST}_{t_s} = \max\{\mathrm{FT}_{t_p} + ct(e_{ps}^{l}), t_p \in \mathrm{pre}(t_s)\} \tag{3}$$

$$\mathrm{FT}_{t_s} = \mathrm{ST}_{t_s} + T_{t_s}^{l} \tag{4}$$

In Eq. (3), $\mathrm{pre}(t_s)$ denotes the collection of tasks that are directly preceding the current task $t_s$. The total amount of time required for completing a whole workflow can be calculated as the time from the beginning of the first task in the observed workflow, until the finish of the last task of the workflow, as given in Eq. (5):

$$T_{\mathrm{total}} = \max\{\mathrm{FT}_{ts}, t_s \in T\} \tag{5}$$

The cost is second important parameter, and it can be separated into the computational cost and the cost required for communication, for both types of resources, edge and cloud. The computational cost, from the task $t_i$ to the task $t_j$ on the server $l$, can be obtained by applying the expression:

$$C_{t_s}^{l} = pp_l \times (\mathrm{FT}_{t_s} - ST_{t_s}), \tag{6}$$

where parameter $pp_l$ denotes the price for processing unit on the server $l$. On the other hand, the communication cost from the task $t_i$ to the task $t_j$ on the server $l$, can be mathematically modeled with Eq. (7):

$$cc(e_{ij}^{l}) = cp_l \times ct(e_{ij}^{l}), \tag{7}$$

where $cp_l$ denotes the price of communication unit on the $l$-th server. The overall cost can then be defined and calculated as the sum of the costs required for computation and communication, as described in Eq. (8):

$$C_{\text{total}} = \sum_{l=1}^{m} \sum_{s=1}^{n} C_{t_s}^{l} + \sum cc(e_{ij}^{l}) \tag{8}$$

At the end, with previously defined calculations of makespan and cost, the objective function in cloud–edge environment can be defined with Eq. (9):

$$\text{obj} = \omega T_{\text{total}} + (1 - \omega) C_{\text{total}}, \tag{9}$$

where $\omega$ denotes weight coefficient, that determines the importance of each objective and it takes a value between 0 and 1. The goal of the objective function specified by Eq. (9) is to minimize the value of obj by balancing between the makespan and cost for a given task of workflow scheduling.

## 4 Proposed Method

Harris hawks optimizer algorithm (HHO) was originally introduced in 2019 by Heidari et al. in [18]. HHO metaheuristics draws the inspiration from the group hunting practice and pursuit tactics of the Harris hawks, which is known as surprise pounce. More information regarding the HHO's background can be retrieved from [18].

The HHO metaheuristics exploration and exploitation phases are motivated by the hawks' exploration for the prey, their special surprise pounce tactic, and various capturing strategies employed by the Harris hawks during the pursuit. The exploration phase of the HHO approach mimics the hawks' process of tracking and detecting the prey. Hawks can wait for hours and observe the target area to identify possible prey. In HHO implementation, each hawk is represented as the candidate solution, while the best solution in every step is observed as the target or close to the optimum. The hawks position themselves on several locations in a random fashion and wait to detect a prey on the basis of two possible strategies. There is an equal probability $q$ for both strategies because hawks in the hunting party position themselves based on the other hawks' locations (to be in the proximity when attacking) and the prey, which can be modeled by using Eq. (10).

$$X(t+1) = \begin{cases} X_{\text{rand}}(t) - r_1 |X_{\text{rand}}(t) - 2r_2 X(t)|, q \geq 0 \\ (X_{\text{best}}(t) - X_m(t)) - r_3(\text{LB} + r_4(\text{UB} - \text{LB})), q < 0.5 \end{cases} \tag{10}$$

where $X(t)$ represents the current hawk (solution) position vector, $X(t+1)$ denotes the solution for the next round t, $X_{\text{best}}(t)$ represents the location of prey (current best solution), coefficients $r_1, r_2, r_3, r_4$, and $q$ are random values within the [0, 1] range

that are being updated in each round, factors $LB$ and $UB$ represent the lower and upper borders of decision variables, $X_{\text{rand}}(t)$ denotes a randomly chosen solution from the current population, and $X_m$ represents the average position of the current solutions population.

The mean location of the current solutions in population ($X_m(t)$) can be obtained by utilizing the following expression:

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \tag{11}$$

where $X_i(t)$ marks the location of $i$-th individual in iteration $t$, while $N$ marks the number solutions in the population.

When transitioning from the exploration process to the exploitation, the HHO approach can switch between several exploitation strategies depending on the prey's available energy (strength of current best solution), which decreases while the prey escapes, as modeled by Eq. (12):

$$E = 2E_0(1 - \frac{t}{T}). \tag{12}$$

where $E$ denotes the prey's energy level that it uses for escaping, $T$ denotes the maximal number of iterations, while $E_0$ represents the initial state of the prey's energy. Parameter $E_0$ switch in random fashion between $(-1, 1)$ in every iteration.

In exploitation phase, the HHO will perform the surprise pounce maneuver to attack the prey which was detected in the previous phase. On the other hand, the prey will try to run away from danger. Depending of the escaping behavior of the rabbit, hawks can employ four possible strategies, which were incorporated in the HHO for modeling the attacking process.

The parameter $r$ is the probability that prey will successfully escape ($r < 0.5$) or not ($r \geq 0.5$) prior to the surprise pounce maneuver. The hunting party of hawks will in both cases encircle the prey and perform hard or soft besiege (determined by the remaining available energy level of the escaping target). The hawks will start with soft besiege, getting closer and closer to the target, and as the prey loses the energy the hawks will intensify the besiege. Parameter E is used for switching between soft and hard besiege process. Soft besiege takes place in case $|E| \geq 0.5$, and hard besiege happens in case $|E| < 0.5$

During the soft besiege, when $|E| \geq 0.5$ and $r \geq 0.5$, the prey still has enough energy for escaping through random jumping. The hawks proceed to encircle the target softly to exhaust it and attack by applying the surprise pounce maneuver, which can be modeled through Eqs. (13) and (14):

$$X(t+1) = \Delta X(t) - E|JX_{\text{best}}(t) - X(t)| \tag{13}$$

$$\Delta X(t) = X_{\text{best}}(t) - X(t) \tag{14}$$

here, $\Delta X(t)$ denotes the distance between the prey's position vector and the current location in the iteration $t$. $J = 2(1 - r_5)$ marks the random jumping strength of the prey during the escape, and $r_5$ is a random value in the interval (0, 1).

Hard besiege, on the other hand, happens when the prey is tired, in other words when $|E| < 0.5$ and $r \geq 0.5$. The hawks in the party proceed to encircle the target hard to finally catch it, and the current positions can be updated by utilizing Eq. (15)

$$X(t+1) = X_{\text{best}}(t) - E|\Delta X(t)| \tag{15}$$

The situation where $|E| \geq 0.5$ and $r < 0.5$ represents the case that the prey's remaining energy is still enough to escape, and soft besiege is still going on prior to the surprise pounce maneuver. This stage is marked with progressive rapid dives around the prey. To simplify the algorithm, it is assumed that the hawks in the group are able to progressively choose the best possible dive toward the best solution. Equation (16) models the hawk's ability to evaluate their next move during the soft besiege:

$$Y = X_{\text{best}}(t) - E|JX_{\text{best}}(t) - X(t)| \tag{16}$$

After evaluation, hawks compare the eventual outcome of dive to the previous one to decide whether it would be a good attempt or not. If they decide that the dive is not reasonable, as the rabbit (best solution) is still moving in misleading steps, hawks additionally commence to dive rapidly and irregularly when approaching the target. To simplify the algorithm, it is assumed that the hawks dive conforming to the LF-based patterns modeled with Eq. (17):

$$Z = Y + S \times \text{LF}(D) \tag{17}$$

where $D$ stands for the problem's dimension, while $S$ denotes the random vector with a size $1 \times D$. The LF is a function that describes the levy flight, which is given with Eq. (18):

$$\text{LF}(x) = 0.01 \times \frac{u \times \sigma}{|v|^{\frac{1}{\beta}}}, \sigma = \left(\frac{\Gamma(1+\beta) \times sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{(\frac{\beta-1}{2})}}\right)^{\frac{1}{\beta}} \tag{18}$$

here, $u$, $v$ are numbers randomly chosen within the interval (0, 1), $\beta$ denotes a fixed value set to 1.5 in the experiments. Finally, the strategy for the soft besiege of the hawks and position updates can be summarized by Eq. (19).

$$X(t+1) = \begin{cases} Y, if\, F(Y) < F(X(t)) \\ Z, if\, F(Z) < F(X(t)) \end{cases} \tag{19}$$

where $Y$ and $Z$ are calculated with Eqs. (16) and (17), respectively.

Finally, in case $|E| < 0.5$ and $r < 0.5$, prey is tired and cannot escape, and hard besiege takes place before surprise pounce maneuver which is used to kill the prey. It

is similar to the soft besiege, however, hawks are closing to the target by decreasing the distance of their average location to the prey, which can be modeled with Eq. (20):

$$X(t+1) = \begin{cases} Y, if\, F(Y) < F(X(t)) \\ Z, if\, F(Z) < F(X(t)) \end{cases} \tag{20}$$

where $Y$ and $Z$ are now calculated with the new Eqs. (21) and (22), respectively:

$$Y = X_{\text{best}}(t) - E|JX_{\text{best}}(t) - X(t)| \tag{21}$$

$$Z = Y + S \times \text{LF}(D) \tag{22}$$

By performing simulations on standard unconstrained instances with the basic HHO, it is noticed that the diversity of population can be enhanced in early phases of the run, while at later stages convergence and the fine searching in the proximity of the current best solution can be accelerated. In this way, original HHO can be improved in terms of the following: if it misses the right portion of the search space, better solutions' diversity will prevent trapping in sub-optimal regions and at later stages, with rational assumption that the right section of the search space is hit, better final solutions' quality can be generated.

Inspired by approach presented in [19], in the basic HHO, this research work has introduced opposition-based learning (OBL) procedure, which can be described as follows: Let $X_j$ denotes $j$th parameter of solution $X$ and the $X_j^o$ represents its opposite number. The opposite number of $j$th parameter of individual $X$ can be calculated as:

$$X_j^o = \text{LB}_j + \text{UB}_j - X_j, \tag{23}$$

where $X_j \in [\text{LB}_j, \text{UB}_j]$ and $\text{UB}_j, \text{LB}_j \in R, \forall j \in 1, 2, 3, \ldots, D$. Notations $\text{UB}_j$ and $\text{LB}_j$ represent lower and upper bound of $j$th parameter, respectively, and $D$ denotes the number of solution dimensions (parameters).

The OBL procedure is implemented in the original HHO in the initialization phase (after the initial population is created) and at the end of each round. For every solution $X_i$ in population $P$, opposite individual $X_i^o$ is generated, and eventually, an opposite population $P^o$ is created. Original and opposite populations are merged together ($P \cup P^o$), solutions in such merged population are sorted in descending order depending of the fitness and first $N$ solutions are chosen as the new population $P$ for the next round.

Inspired by included changes in the original HHO, the proposed approach HHO is named with OBL procedure (HHOOBLP). The pseudocode for the proposed HHOOBLP is summarized in Algorithm 1.

**Algorithm 1** Proposed HHOOBLP pseudo-code

```
Inputs: The population size N and maximum amount of rounds T
Output: The location of the best solution and its fitness value
Generate the initial population P_INT of random solutions X_i, (i = 1, 2, 3, ... N)
Generate opposite initial population P^o_INT, perform P_INT ∪ P^o_INT and select N best solutions
while exit criteria has not been fulfilled do
   Determine the fitness values for each hawk
   Set X_best as the rabbit's position (best location)
   for each solution X_i do
      Update the starting energy E_0 and jumping strength J : E_0 = 2rand() − 1, J = 2(1 − rand())
      Update the value E by using Eq. (12)
      if |E| ≥ 1 then
         Exploration stage
         Update values of the location vector according to the Eq. (10)
      end if
      if |E| < 1 then
         Exploitation stage
         if r ≥ 0.5 and |E| ≥ 0.5 then
            Soft besiege phase
            Update values of the location vector according to the Eq. (13)
         else if r ≥ 0.5 and |E| < 0.5 then
            Hard besiege phase
            Update values of the location vector according to the Eq. (15)
         else if r < 0.5 and |E| ≥ 0.5 then
            Soft besiege phase with rapid diving
            Update values of the location vector according to the Eq. (19)
         else if r < 0.5 and |E| < 0.5 then
            Hard besiege phase with rapid diving
            Update values of the location vector according to the Eq. (20)
         end if
      end if
   end for
   Generate population P^o, perform P ∪ P^o and select N best solutions
end while
Return X_best
```

## 5 Experiments

First, this section provides the details of simulation environment and experimental setup and then the empirical simulations' results are shown and comparative analysis has been performed between our proposed HHOOBLP and original HHO along with other outstanding heuristics and metaheuristics.

### *5.1 Simulation Environment Setup*

The same environment and experimental setup is used as in [2], because this research work intends to conduct fair comparative evaluation with other sophisticated approaches that are validated against the same workflow instance that were shown in this paper.

The workflow simulator which is utilized in all conducted experiments is WorkflowSim-1.0, created and maintained by the Pegasus WMS group at the University of Southern California, and published under the GNU General Public License [20].

This simulator is chosen due to its ability to provide an environment, which is the excellent approximation of real distributed system. The simulations were executed on the system with Intel® 7700 i7 3.6 GHz, 16 GB of RAM and with Windows 10 OS.

The HHOOBLP is implemented in Java and tested it for the five workflow models with small number of tasks, which are developed by Pegasus group: CyberShake, Epigenomics, Inspiral, Montage, and Sipht with 30, 24, 30, 25, and 30 task nodes, respectively. Moreover, the proposed research work has utilized *Montage* 100 workflow to determine the proper value of the weight parameter $\omega$ for the objective function. All workflows simulate different real applications and follow typical DAG structure, which is visualized in Fig. 1. Properties of models are given in Table 1. Moreover, to measure performance improvements of proposed HHOOBLP over the original HHO, also the basic HHO is implemented and tested. Both approaches were integrated in the WorkflowSim environment. As in [2], 6 cloud and 4 edge servers are used in simulations. The parameters of the cloud/edge servers, with different processing rates and communication properties, are presented in Table 2.

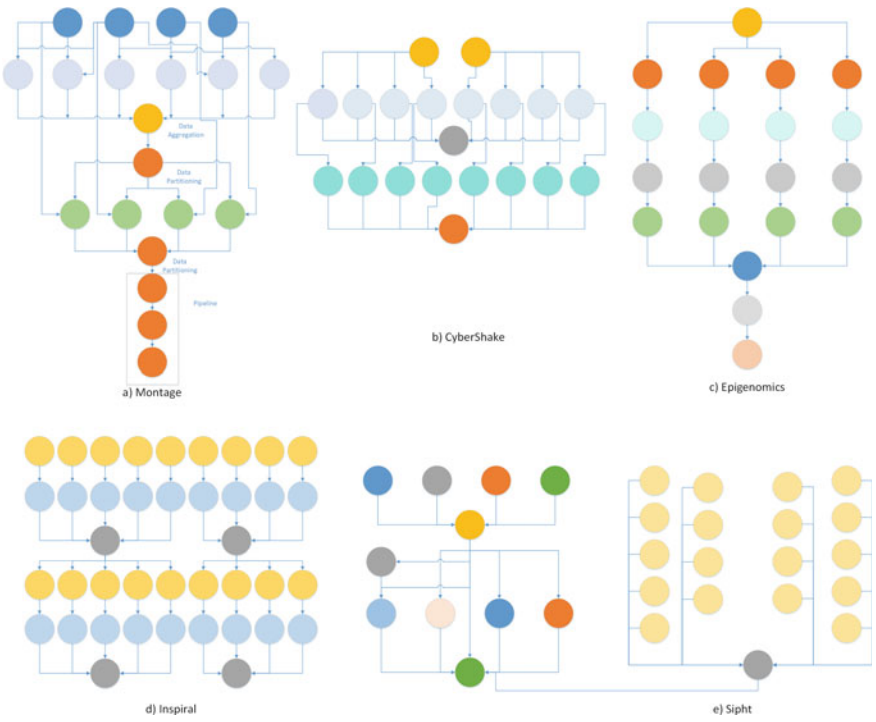



**Fig. 1** Workflow models used in the experiments

**Table 1** Properties of real-world DAGs

| DAG | No. of nodes | No. of edges | Average data size (MB) | Average task runtime |
|---|---|---|---|---|
| CyberShake_30 | 30 | 112 | 747.48 | 23.77 |
| Epigenomics_24 | 24 | 75 | 116.20 | 681.54 |
| Inspiral_30 | 30 | 95 | 9.00 | 206.78 |
| Montage_25 | 25 | 95 | 3.43 | 8.44 |
| Sipht_30 | 30 | 91 | 7.73 | 178.92 |
| Montage_100 | 100 | 433 | 3.23 | 10.58 |

**Table 2** Parameters for the cloud and edge computing resources

| Parameter | Servers | Proc. rate (MIPS) | Proc. cost (per time units) | Bandwidth (Mbps) | Com. cost (per time units) |
|---|---|---|---|---|---|
| Cloud servers | 0 | 5000 | 0.5 | 800 | 0.5 |
| | 1 | 5000 | 0.5 | 500 | 0.4 |
| | 2 | 3500 | 0.4 | 800 | 0.5 |
| | 3 | 3500 | 0.4 | 500 | 0.4 |
| | 4 | 2500 | 0.3 | 800 | 0.5 |
| | 5 | 2500 | 0.3 | 500 | 0.4 |
| Edge servers | 6 | 1500 | 0.2 | 1500 | 0.7 |
| | 7 | 1500 | 0.2 | 1000 | 0.6 |
| | 8 | 1000 | 0.1 | 1500 | 0.7 |
| | 9 | 1000 | 0.1 | 1000 | 0.6 |

## *5.2 Comparative Analysis and Discussion*

This subsection exhibits the summed results of the comparative evaluation of the proposed HHOOBLP approach with other metaheuristics and heuristics which are considered to be state of the art. The experimental environment, experimental conditions, and the tested workflows were the same for all observed approaches. Throughout conducted experiments, the practice has been followed from [2]. The proposed HHOOBLP is evaluated by comparing it to the original HHO version, directional and non-localconvergent PSO (DNCPSO) [2], and also other heuristics (HEFT and MIN-MIN), as classic scheduling methods and metaheuristics (original PSO and GA), as algorithms that perform dynamic scheduling. For the purpose of performing objective comparative evaluations with opponent approach, the original HHO has been tested and proposed HHOOBLP with 300 iterations ($T = 30$) and with population of 50 solutions ($N$=50). Moreover, for each experiment instance the algorithms are executed in 10 independent runs and average results are reported.

Weight coefficient $\omega$ of objective function, given in Eq. (9) for all algorithms included in comparative analysis was set to 0.8, because with this settings, best trade-off between makespan and cost can be established. The simulations are conducted with different $\omega$ values within the range of $[0.1, 0.9]$ with step length of 0.1 for *Montage* workflow with 100 task nodes, as in [2]. Based on obtained results, this research work has been concluded that by changing $\omega$, makespan has higher variance than the cost, which further means that the makespan has greater influence on the objective. Visual representation of simulation results of makespan, cost, and objective metrics for variable $\omega$ of proposed HHOOBLP is given in Fig. 2.

Simulation results for other methods which were covered in the comparative analysis have been obtained from [2]. The graphical representation of the conducted evaluation for makespan, cost, and combined objective for all observed approaches is shown in Fig. 3. The smaller values of the makespan, cost, and combined objective indicate better performances of the algorithm. As it can be clearly observed from the simulation results given in Fig. 3, on average, by taking into the account all observed test instances, the proposed HHOOBLP achieves the best performances for all three evaluated metrics (makespan, cost, and comined objective). The original PSO established the worse performance, as expected, since it does not implement efficient exploration mechanism and can be trapped in sub-optimal region. The GA performed better than the original PSO for all indicators, as it can be easily adapted for discrete optimization problems by using proper encoding strategy. Original HHO performed better than the PSO, however, like PSO, it can also get trapped into the sub-optimal regions of the search space, but due to the more efficient exploitation procedure, it obtained better results.

The DNCPSO has proven as an efficient metaheuristics, that overcomes issues exposed by the original PSO. Nevertheless, our proposed HHOOBLP metaheuristics has proven to be even more efficient. For instance, when *Cybershake 30* test is observed, the DNCPSO achieved slightly lower value of cost, however, HHOOBLP established better result for the makespan, and the objective value. Similar pattern can be seen on *Montage 25* and *Sipht 30* instances, where DNCPSO obtained slightly lower values for cost, but HHOOBLP outperformed it in terms of makespan and

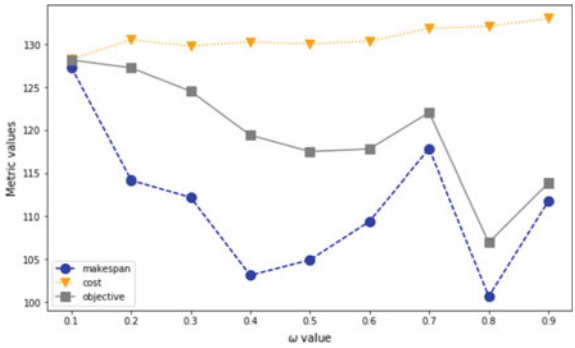



**Fig. 2** Makespan, cost, and objective for different values of $\omega$ for *Montage* 100 workflow of proposed HHOOBLP

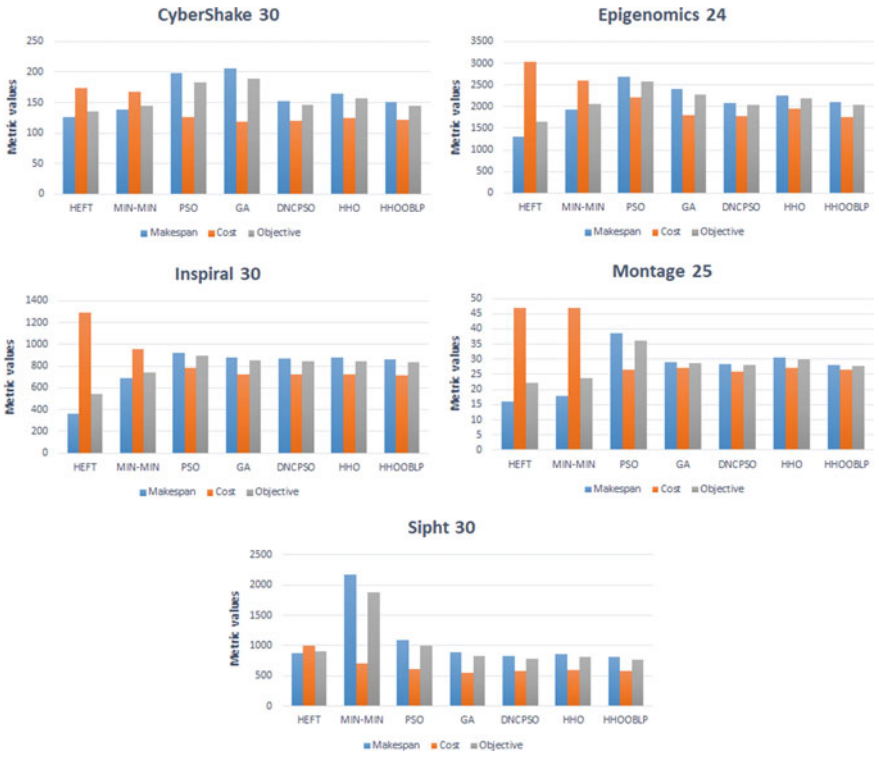



**Fig. 3** Simulation results—comparison between HHOOBLP, original HHO, DNCPSO, and other approaches for datasets with smaller number of task nodes

overall objective value. In case of *Inspiral 30* instance, the HHOOBLP outperformed DNCPSO both in cost and makespan, consequently obtaining better objective value as well. Only on the instance of *Epigenomics 24* DNCPSO achieved better values than HHOOBLP, both in terms of makespan and cost and objective. The original HHO metaheuristics also achieved solid results, with the performances just behind the DNCPSO.

On the other hand, heuristic approach represented in conducted experiments with HEFT and MIN-MIN algorithms is oriented to minimize the makespan, due to the static scheduling process. Consequentially, makespan for HEFT and MIN-MIN algorithms is in general very short. On the other hand, the cost of these algorithms is very high. The *Sipht 30* test simulation is the only exception, where MIN-MIN allocates some tasks with heavy computational load to the units with low processing power, therefore making the makespan much higher. In general, as the chosen value $theta = 0.8$ favors makespan, which in turn has more influence on the combined objective than the cost, value of objective for both observed heuristics is quite low.

However, even with the good makespan values, and consequently good objective value, the cost of HEFT and MIN-MIN approaches is very high. This is typically not

acceptable for the real-life workflow applications, where cost plays an important role. On the other hand, dynamic scheduling provided by the metaheuristics approaches gives much better balance between the makespan and cost, and therefore they are much more suitable for the real-life applications in the real cloud–edge environments.

## 6 Conclusion

This research work has proposed and implemented an improved HHO approach, that enhances original HHO by including OBL procedure for solving the problem of cloud–edge workflow scheduling. Based on simulation results, it was proven that the proposed HHOOBLP obtains better results than the original HHO and other sophisticated heuristics and metaheuristics for workflow scheduling challenge in cloud–edge systems.

According to the accomplished results, main contributions of proposed research are enhancements of the basic HHO and improvements of the makespan and cost for tackling cloud–edge workflow scheduling issue. As part of our future work in this domain, our intent is to carry on the research with the promising HHO approach and tackle different problems in cloud–edge environments.

## References

1. Shiliang L, Lianglun C, Bin R (2014) Practical swarm optimization based fault-tolerance algorithm for the internet of things. KSII Trans Internet Inf Syst 8(4):1178–1191
2. Xie Y, Zhu Y, Wang Y, Cheng Y, Xu R, Sani AS, Yuan D, Yang Y (2019) A novel directional and non-local-convergent particle swarm optimization based workflow scheduling in cloud-edge environment. Future Gener Comput Syst 97:361–378
3. Thanh Dat D, Doan H (2017) Fbrc: optimization of task scheduling in fog-based region and cloud. In: IEEE Trustcom/BigDataSE/ICESS, vol 2017, pp 1109–1114
4. Wang H, Wang Y (2018) Maximizing reliability and performance with reliability-driven task scheduling in heterogeneous distributed computing systems. J Ambient Intell Humanized Comput
5. Wang T, Liu Z, Chen Y, Xu Y, Dai X (2014) Load balancing task scheduling based on genetic algorithm in cloud computing. In: 2014 IEEE 12th international conference on dependable, autonomic and secure computing, pp 146–152
6. Zhan Z-H, Zhang G-Y, Gong Y-J, Zhang J (2014) Load balance aware genetic algorithm for task scheduling in cloud computing. In: Dick G, Browne WN, Whigham P, Zhang M, Bui LT, Ishibuchi H, Jin Y, Li X, Shi Y, Singh P, Tan KC, Tang K (eds) Simulated evolution and learning. Springer International Publishing, Cham, pp 644–655
7. Yang X-S (2014) Swarm intelligence based algorithms: a critical analysis. Evol Intell 7:17–28
8. Strumberger I, Bacanin N, Tuba M (2017) Enhanced firefly algorithm for constrained numerical optimization, ieee congress on evolutionary computation. In: Proceedings of the IEEE international congress on evolutionary computation (CEC 2017), pp 2120–2127
9. Tuba M, Bacanin N (2014) Improved seeker optimization algorithm hybridized with firefly algorithm for constrained optimization problems. Neurocomputing 2(143):197–207

10. Bacanin N, Tuba M (2012) Artificial bee colony (ABC) algorithm for constrained optimization improved with genetic operators. Stud Inf Control 21:137–146
11. Bacanin N, Tuba M (2014) Firefly algorithm for cardinality constrained mean-variance portfolio optimization problem with entropy diversity constraint. Sci World J Special issue Computational Intelligence and Metaheuristic Algorithms with Applications 2014(Article ID 721521):16
12. Strumberger I, Tuba E, Bacanin N, Beko M, Tuba M (2018) Wireless sensor network localization problem by hybridized moth search algorithm. In: 2018 14th International wireless communications mobile computing conference (IWCMC), pp 316–321
13. Sagnika S, Bilgaiyan S, Mishra BSP (2018) Workflow scheduling in cloud computing environment using bat algorithm. In: Proceedings of first international conference on smart system, innovations and computing. Springer, pp 149–163
14. Kumar M, Sharma S (2018) Pso-cogent: cost and energy efficient scheduling in cloud environment with deadline constraint. Sustain Comput Inf Syst 19:147–164
15. Agarwal M, Srivastava GMS (2018) A cuckoo search algorithm-based task scheduling in cloud computing. In: Bhatia SK, Mishra KK, Tiwari S, Singh VK (eds) Advances in computer and computational sciences. Springer Singapore, Singapore, pp 293–299
16. Strumberger I, Tuba M, Bacanin N, Tuba E (2019) Cloudlet scheduling by hybridized monarch butterfly optimization algorithm. J Sens Actuator Netw 8(3):44
17. Strumberger I, Bacanin N, Tuba M, Tuba E (2019) Resource scheduling in cloud computing based on a hybridized whale optimization algorithm. Appl Sci 9(22):4893
18. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: algorithm and applications. Future Gener Comput Syst 97:849–872
19. Abd Elaziz M, Oliva D (2018) Parameter estimation of solar cells diode models by an improved opposition-based whale optimization algorithm. Energy Convers Manage 1(171):1843–1859
20. Chen W, Deelman E (2012) Workflowsim: a toolkit for simulating scientific workflows in distributed environments. In: 2012 IEEE 8th international conference on E-science. IEEE, pp 1–8

# Generation of Random Binary Sequence Using Adaptive Row–Column Approach and Synthetic Color Image

**C. Manikandan, N. Raju, K. Sai Siva Satwik, M. Chandrasekar, and V. Elamaran**

**Abstract** The security of a communication model is defined by the strength of the encryption and key generation algorithm. To ensure security, in block cipher techniques, a complex computation process is used for encryption. But, in stream cipher techniques, complex key generation techniques are used. This paper proposes a novel and complex key generation algorithm for stream cipher techniques that generate 15,72,864-bit key from a synthetic color image using a pattern-based bit extraction technique. A segmented form of a generated key can be used as a dynamic key in block cipher techniques. The proposed algorithm randomly uses eight patterns to extract bits from a synthetic color image to generate the key. The generated key's randomness is tested using the NIST statistical analysis tool and compared with the keys generated from existing techniques. The key length and keyspace are compared with existing methods.

**Keywords** Key generation · LSB extraction · NIST statistical analysis

## 1 Introduction

The growing use of digital communication like the Internet, Wi-Fi, Li-Fi, Internet of Things (IoT) devices, and so on demands improving security standards. Many encryption techniques are implemented for secure data transmission [1]. These encryption techniques typically use secret strings consisting of alphabets, numbers, characters, or binary digits to encrypt the sensitive data. These strings are called "keys," and encrypted data is called "ciphertext." The ciphertext is communicated to other users publicly, and keys are shared secretly through secure means. On the other side, users having a decryption algorithm receive secret keys and ciphertext to get plain text [2]. The encryption algorithm's strength is crucial in this process because a robust algorithm uses strong keys and generates ciphertext that is hard to decrypt by any

C. Manikandan · N. Raju (✉) · K. Sai Siva Satwik · M. Chandrasekar · V. Elamaran
School of EEE, SASTRA Deemed to be University, Thanjavur, Tamil Nadu 613401, India
e-mail: raju@ece.sastra.edu



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_10

hackers without the secret keys. Some of the more powerful and standard encryption algorithms are data encryption standard (DES), advanced encryption standard (AES), Rivest–Shamir–Adleman (RSA), Rivest Cipher (RC4), blowfish, and so on. All these algorithms use keys varying from 32 bits to 4096 bits [3]. Although the algorithms mentioned above are powerful for encryption, the key generation process in these algorithms can sometimes be decoded by hackers with excellent knowledge and access to the algorithm [4]. Therefore, to improve the strength of the key generation algorithm, various techniques had been implemented. For this paper's scope, only the methods that involve key generation from digital images are considered. Digital image key generation techniques were extracting pixel values or image properties and converting them into keys [5, 6]. The method proposed in [7] used LSB extraction and the absolute difference computation to generate the key. In [8], the histogram of the image is calculated, and the key is selected based on the peak value. In [9], a chaotic map-based key generation technique was proposed, pixels are selected based on the generated chaotic map, and LSB values are XORed for key generation. The method in [10] uses an alphabetical trie approach that matches the chosen character set to the image by dividing the image into blocks row-wise. The mean of the matched blocks is calculated and converted into a binary form to generate the key. LFSR-based key generation technique is proposed in [11], which selects bits from each pixel by mapping the generated 3-bit LFSR sequence. The algorithm proposed in [12] generates dynamic nonlinear keys using logistics and piecewise chaotic maps. These algorithms create keys of acceptable length and in less time. Many of these algorithms generate larger keys. But the logic used for these algorithms were simple and easy to understand from a hacker's perspective. Hackers can easily recreate the key. The other limitation for algorithms proposed in [5] is the memory space requirement. In [13, 14], variable size dynamic key generation techniques were proposed. In [5], various pictures in the image database generate a dynamic key by quickly changing the images. This requires more storage space to store images. In [12], the active key's keyspace is around $2^{256,}$ which is acceptable but can be cracked in very little time using quantum computing algorithms. Due to the advancement in quantum computing technology, the computational speed and time to crack an algorithm have significantly improved. Many more robust algorithms like DES and AES can be cracked in a few hours using quantum computers. Therefore, there is a huge necessity to enhance the key generation algorithm's strength and length of the key [15]. The primary goal is to design an extensive and complex key generation algorithm. It can generate larger keys, increasing the time taken to crack the key generation algorithm. This work also meets the requirements of a dynamic key generation algorithm. This paper proposes a new pattern-based key generation algorithm from a synthetic color image. The proposed work in [16, 17] uses eight different patterns, namely Straight Up (SU), Straight Down (SD), Straight Forward (SF), Straight Backward (SB), Flipped Up (FU), Flipped Down (FD), Flipped Forward (FF), and Flipped Backward (FB), for embedding binary data into LSB of the pixels. In this paper, same eight patterns are used to extract the pixels' binary values from LSB to MSB. The extracted bits from each plane are stored in buffers. A total of 3

keys are generated from each plane of the image, and those three keys are concatenated to create one final key. The proposed work is arranged as follows. Section 2 provides methodology, a flowchart, and a step-by-step process of the key generation algorithm. Section 3 analyzes the key's randomness, the complexity of the key generation, and the proposed algorithm's comparison with the existing algorithms. Section 4 concludes the proposed work based on the results and analysis.

## 2 Methodology

$N \times N$ synthetic images, namely red plane image, green plane image, and blue plane image, were generated using a pseudo-random generator. Each pixel in these images is represented by 8-bit values varying from 0 to 255. These three images are combined, respectively, to form a synthetic color image, as shown in Fig. 1. This synthetic color image is used for the proposed algorithm.

The proposed key generation algorithm uses eight different patterns mentioned in [16, 17] for extracting the bit values from each pixel of the selected block, as shown in Fig. 2. The numbers in Fig. 2 indicate how the columns/rows of the pixels are selected for bit extraction with 1 being the first column/row and 8 being the last column/row. The arrows in Fig. 2 indicate the direction in which the pixels in that particular column/row are selected for bit extraction. For example, in Fig. 2a, the first column is selected, and pixels are arranged from bottom to top based on the arrow

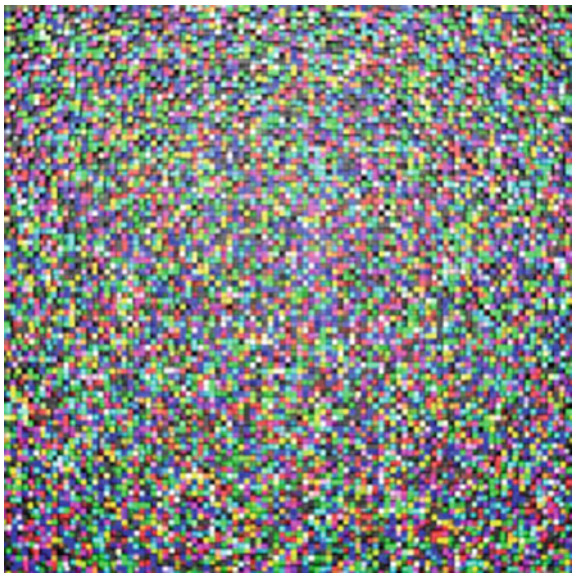


**Fig. 1** Synthetic color image

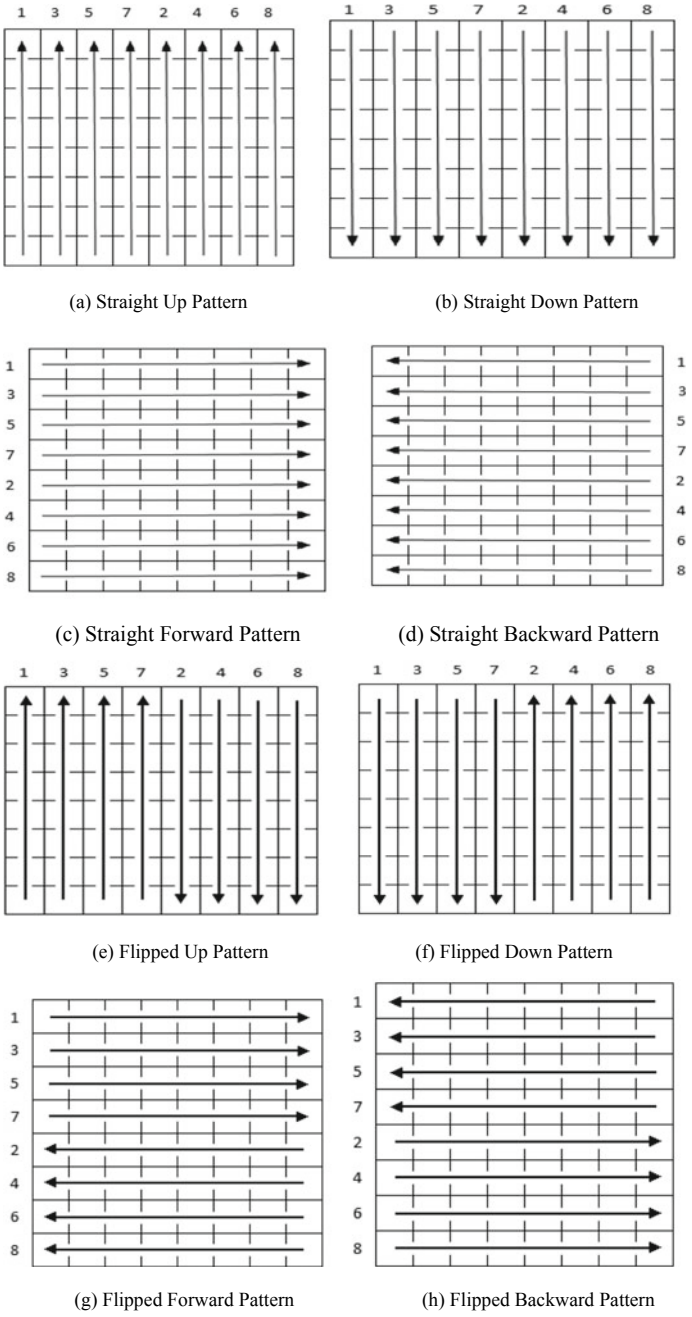



**Fig. 2** Bit extraction patterns

in an array. The next fifth column is selected, and pixels are arranged from bottom to top. The same process is repeated for all eight columns, and pixels are organized based on arrows. The bit values are extracted from the arranged pixels, as proposed in the algorithm below.

The flowchart representation of the proposed algorithm is given in Fig. 3. The step-by-step process of key generation using the proposed algorithm is given below.

Step 1: Read the synthetic color image of size N × M.

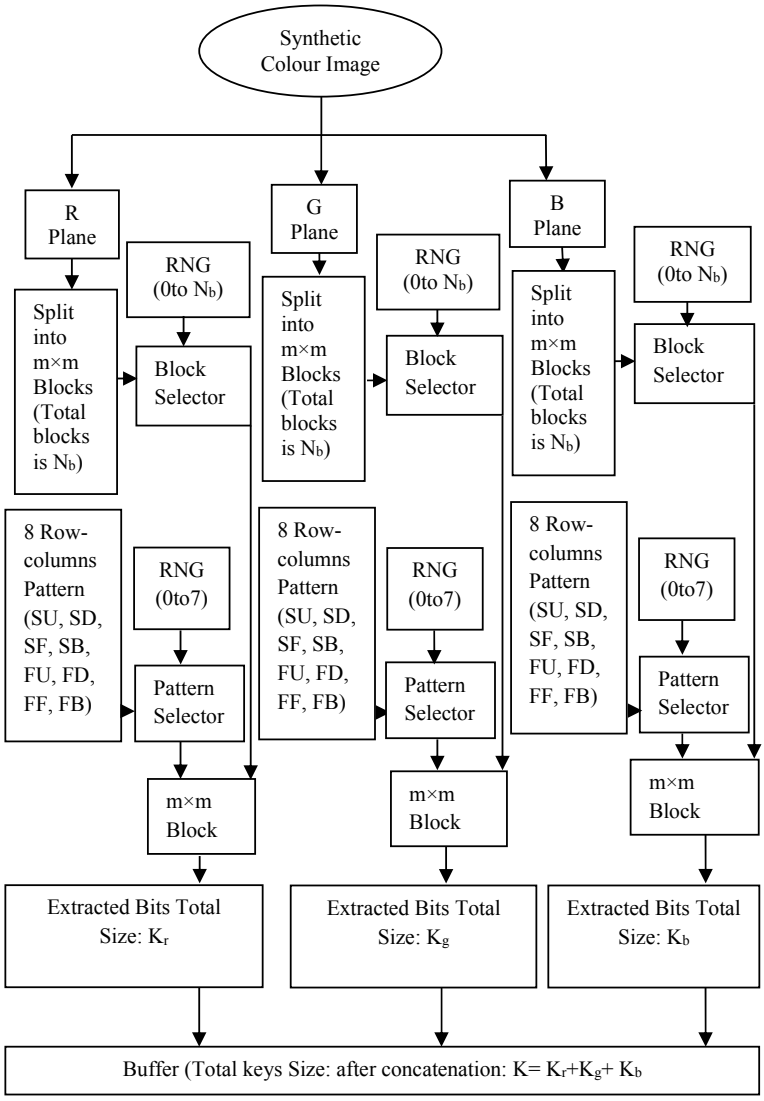



**Fig. 3** Flowchart representation of the proposed algorithm

Step 2: Divide the color image into three planes, namely red, green, and blue.

Step 3: Select the red layer and divide it into m × m blocks. This gives a total of blocks $N_b = (N \times M) \div (m \times m)$.

Step 4: A random number generator that generates numbers from 0 to $N_b$ is randomly selected. Each block is chosen randomly at a time using the random generator.

Step 5: Another random number generator that varies from 0 to 7 is created.

Step 6: Based on the number generated by the second random generator, one among the eight approaches (SU, SD, SF, SB, FU, FD, FF, and FB) is selected.

Step 7: The selected approach is applied for the selected block, and every bit of the pixels in that block is selected from LSB to MSB one at a time and stored as keys. Therefore, for one m × m block, a total of 8 keys of length m × m bits are generated.

Step 8: Concatenate the 8 keys generated from step 7 to create one key of length $K_t = m \times m \times 8$ bits.

Step 9: Repeat this process from step 4 to step 8 for the whole red layer until all the blocks are selected randomly. This gives $N_b$ keys, each of length $K_t$ bits.

Step 10: Concatenate $N_b$ keys generated from step 9 to create a key of length $K_r = N_b \times K_t$ bits.

Step 11: Simultaneously, the green and blue layers are selected, and the process is repeated from step 3 to step 10. Hence, for one cycle, a total of 3 keys $K_r$, $K_g$, and $K_b$, each with a key length of $N_b \times K_t$ bits are generated, respectively.

Step 12: Concatenate keys from red, green, and blue planes to create a final key of length $K = K_r + K_g + K_b$ bits.

The proposed complex algorithm is implemented for a 256 × 256 synthetic color image. The algorithm gives the key's length, as shown in Eqs. (1) and (2)

$$K = \frac{(N \times M)}{(m \times m)} \times (m \times m) \times 8 + \frac{(N \times M)}{(m \times m)} \times (m \times m) \times 8 + \frac{(N \times M)}{(m \times m)} \times (m \times m) \times 8 \tag{1}$$

$$K = K_r + K_g + K_b \tag{2}$$

Using these equations, the total length of the key generated using a 256 × 256 synthetic color image is $K = 1{,}572{,}864$ bits.

## 3 Results and Discussion

The proposed algorithm generates a key length of 1,572,864 bits from the synthetic color image. The strength of the key places a vital role in determining the security of an encryption algorithm. The key's strength is determined by the randomness of 1s and 0s in the key. The randomness of the key is tested using the NIST statistical analysis tool. NIST statistical analysis performs various tests like frequency, block frequency, cumulative sums, runs, longest runs, rank, FFT, non-overlapping template, approximate entropy, serial and linear complexity. The results generated from the NIST statistical analysis tool are given in Table 1. It is observed that the generated key has passed the entire randomness tests with the probability value greater than the benchmark (>0.01).

The strength of the key generation algorithm must also be considered because any hacker with the knowledge of the key generation algorithm can generate his own key. If the key generation algorithm isn't strong enough, then there is a high probability of the hacker's key to match the original key. The strength of the key generation algorithm can be determined by estimating the complexity of key generation. If the hacker doesn't know the key generation's algorithm, it takes $2^{1,572,864}$ ways to identify the key. If the attacker knew the proposed algorithm used for key generation, the complexity of generating the original key is calculated. Each $8 \times 8$ block is selected randomly using a random number generator. Therefore, the total number of ways to select all the blocks is 1024! Since there are 8 different patterns were used for each block, bit extraction can be done from each block in 8 ways. Therefore, the total ways to takes to estimate the key are $8^{1024} \times (1024)!$.

The proposed algorithm is compared with the existing algorithms like LSB histogram, logistic map, LFSR key generation for key length, and keyspace are shown in Table 2. All these algorithms are implemented using MATLAB, and results

**Table 1** Results from the NIST statistical analysis tool

| Test | Probability | Result |
|---|---|---|
| Frequency | 0.911413 | Pass |
| Block frequency | 0.534146 | Pass |
| Cumulative sums | 0.739918 | Pass |
| Runs | 0.739918 | Pass |
| Longest run | 0.213309 | Pass |
| Rank | 0.534146 | Pass |
| FFT | 0.350485 | Pass |
| Non-overlapping template | 0.911413 | Pass |
| Approximate entropy | 0.213309 | Pass |
| Serial | 0.739918 | Pass |
| Linear complexity | 0.012043 | Pass |

**Table 2** Comparison of the proposed algorithm with existing algorithms

| Algorithm | Key size (bits) | Key space |
|---|---|---|
| LSB histogram | 1776 | $2^{1776}$ |
| Logistic map | 65,536 | $2^{65,536}$ |
| LFSR | 196,608 | $2^{196,608}$ |
| Proposed algorithm | 1,572,864 | $2^{1,572,864}$ |


**Fig. 4** Application of the proposed framework

are generated for ideal conditions. It is observed from Table 2 that the proposed algorithm has a key length of 1,572,864 and keyspace of $2^{1,572,864.}$ which are greater than the existing algorithms. The keyspace is calculated using the standard formula $2^k$, where $k$ is the length of the key.

The proposed key generation algorithm's application is demonstrated with stream ciphers, as shown in Fig. 4. The plaint text in ASCII format is converted into a binary stream. Each bit of the plain binary stream is XORed with each bit of the key generated using the proposed algorithm, and binary ciphertext is generated. The binary ciphertext is communicated to the receiver. The binary cipher stream is XORed with the received key to generate plain binary text at the receiver side. Finally, the plain binary text is converted into ASCII format, and the message is read. This application is implemented using MATLAB GUIDE, and the obtained results are shown in Fig. 4.

The proposed key generation algorithm can also be used for block ciphers like DES, where the total length of the key is divided into individual blocks of 64-bit or 128-bit [15, 16]. Each block of the key is used to encrypt each 64-bit or 128-bit message block to generate the ciphertext. On the receiver side, Each 64-bit or 128-bit cipher block is decrypted using each 64-bit or 128-bit key block, and plain text is recovered. Table 3 shows how many attacks are required to get plaintext from the ciphertext if the proposed key is used for the existing stream and block cipher

**Table 3** Robustness against malicious attacks

| Encryption algorithm | Key size (bits) | No. of attacks to decrypt |
|---|---|---|
| Stream (XOR) | 64 (standard) | $2^{64}$ |
| | 1,572,864 | $2^{1,572,864}$ |
| DES | 64 (standard) | $2^{56}$ |
| | 1,572,864 | $24{,}576 \times 2^{56}$ |

**Table 4** Key length for various synthetic color image resolutions

| Resolution | Size | Key length (bits) |
|---|---|---|
| 480p | $640 \times 480$ | 7,372,800 |
| 720p | $1280 \times 720$ | 22,118,400 |
| 1080p | $1920 \times 1080$ | 49,766,400 |
| 4K | $3840 \times 2160$ | 199,065,600 |

encryption algorithm. For, block encryption 1,572,864 bit key is divided into 64-bit keys blocks. Therefore, a total of 24,576 keys each of length 64-bit are created. These keys are used to encrypt the incoming 64-bit message blocks.

The generated key can be used to encrypt various data like text, image, video, file, and sensor output represented in binary form. The length of the key generated using the proposed algorithm can be changed by changing the synthetic color image's resolution. Table 4 provides various lengths of the key generated using different synthetic images of standard resolutions. These resolutions vary from 480p to 4K, which is in practical use. The size of the key is calculated using Eq. 1. Table 4 shows that a maximum key length of 199,065,600 bits can be generated using the proposed algorithm when a $3840 \times 2160$ synthetic color image is used.

Although the proposed algorithm has all the advantages mentioned above over existing algorithms, there are few limitations to which the proposed algorithm may not be applicable. The proposed algorithm requires the incoming data to be as large as the generated key to avoid repetition and redundancy. The proposed algorithm cannot be used in situations where sensitive data is minimal, typically in kilobytes. The proposed algorithm is not a lightweight algorithm and sometimes more challenging for the developers to understand.

## 4 Conclusion

This paper presents a new key generation algorithm using a synthetic color image. The proposed algorithm uses a pattern-based bit extraction technique for key generation. There are eight different patterns were used for the extraction of bits from LSB to MSB. One pattern is selected randomly and applied for a randomly selected block

to extract bits from LSB to MSB. These extracted bits are arranged to create a key of length 1,572,864 bits. The randomness of the generated key is tested using the NIST statistical analysis tool, thus proving its strength. The proposed algorithm is compared with the existing algorithms and observed that the proposed algorithm generates a larger size key. The complexity estimation of the proposed algorithm shows that it takes $8^{1024} \times (1024)!$ Ways to break the algorithm. The results concluded that the proposed key generation algorithm suitable for a dynamic key-dependent-based stream and block cipher algorithm. The proposed key generation algorithm is applicable for massive size data and is not for small size data. In the future, the key generation algorithm can be further modified to generate keys from various image types like medical images, bitmap images, satellite images, aerial images, and so on. The randomness of the generated key using the proposed algorithm can be further improved using various computational techniques like bit inversion, s-box, binary shift, and bit rotation.

## References

1. Shiu YS, Chang SY, Wu HC, Huang SCH, Chen HH (2011) Physical layer security in wireless networks: a tutorial. IEEE Wirel Commun 18(2):66–74
2. Liu Y, Chen HH, Wang L (2016) Physical layer security for next-generation wireless networks: theories, technologies, and challenges. IEEE Commun Surv Tutor 19(1):347–376
3. Singh G (2013) A study of encryption algorithms (RSA, DES, 3DES, and AES) for information security. Int J Comput Appl 67(19)
4. Erickson J (2008) Hacking: the art of exploitation. No starch press
5. Santhi B, Ravichandran KS, Arun AP, Chakkarapani L (2012) A novel cryptographic key generation method using image features. Res J Inf Technol 4(2):88–92
6. Manikandan C, Neelamegam P, Kumar R, Babu V, Satwikkommi S (2019) Design of Secure and Reliable MU-MIMO Transceiver System for Vehicular Networks. Int J Comput Netw Commun 11:15–32
7. Barhoom TS, Abusilmiyeh ZM (2013) A novel cryptography method based on image for key generation. In: Palestinian international conference on information and communication technology. IEEE, Gaza, Palestinian, pp 71–76
8. Riddhi N, Gamit N (2015) An efficient algorithm for dynamic key generation for image encryption. In: International conference on computer, communication, and control. IEEE, Indore, India, pp 1–5
9. Manikandan C, Kumar SR, Nikhith K, Gayathri MS, Neelamegam P (2019) Chaotic map based key generation and realistic power allocation technique for secure MU-MIMO wireless system. In: International conference on applications and techniques in information security, Springer, Singapore, pp 142–155
10. Manikandan G, Ramakrishnan S, Rajaram R, Venkatesh V (2013) An image-based key generation for symmetric key cryptography. Int J Eng Technol 5(3):2807–2810
11. Chinnusamy M, Sidharthan RK, Sivanandam V, Kommi SSS, Mallari Rao C, Periasamy N (2019) Optimal tracking of QR inspired LEA using particle filter for secured visual MIMO communication based vehicular network. Photonics 6(93):1–24
12. Jawad LM, Sulong G (2015) A novel dynamic secret key generation for an efficient image encryption algorithm. Mod Appl Sci 9(13):85–97
13. Yao J, Kang H (2011) FPGA implementation of dynamic key management for des encryption algorithm. In: International conference on electronic & mechanical engineering and information technology. IEEE, Harbin, China, pp 4795–4798

14. Singh AK, Varshney S (2014) Enhanced data encryption standard using variable size key (128 N Bits) and 96 bit subkey. Int J Comput Appl 98(8):11–14
15. Chen L, Jordan S, Liu YK, Moody D, Peralta R, Perlner R, Smith-Tone D (2016) NIST: report on post-quantum cryptography. NIST, Tech. Rep.
16. Satwik KSS, Manikandan C, Elamaran V, Narasimhan K, Raju N (2017) An adaptive row-column least significant bit inlay approach for image steganography. Biomed Res 28(22):10216–10222
17. Manikandan C, Rakesh Kumar S, Sai Siva Satwik K, Neelamegam P, Narasimhan K, Raju N (2018) An integrated object tracking and covert visual MIMO communication service for museum security system using single vision sensor. Appl Sci 8(10):1–25

# Blockchain: Application Domains, Research Issues and Challenges


**Dipankar Debnath and Sarat Kr. Chettri**

**Abstract** Blockchain technologies offer an innovative approach to data storage, transaction execution, process automation, and confidence-building in an open and distributed environment. A wide range of blockchain-based applications have been developed and deployed since its inception, and the convergence of blockchain with other state-of-the-art technologies has been on the rise. However, the issues and challenges associated with it are also rising with the advancement of blockchain technologies and their applications. This paper provides an overview of blockchain and reviews the typical domains of blockchain application discussing the open issues and challenges. Besides, this article reviews the recent advances in addressing different blockchain issues and points out the research directions that help to develop technological innovations for future blockchain systems.

**Keywords** Blockchain · Consensus algorithms · Security and privacy · IoT · Smart contract · Cryptocurrency · Issues and challenges

## 1 Introduction

In simple terms, a blockchain can be defined as a time-stamped series of data records that are immutable and managed by a group of computers without any central authority. Such data records are securely stored as blocks, which are linked to each other using the ideologies of cryptography. Blockchain technology is seen as a recent breakthrough in an open-networked system where computation is secure and transparent and where everyone involved is accountable for their actions without any ownership by a single entity. The technology which was originally devised for digital currency has now found its potential uses in various other areas. The global annual

D. Debnath (✉)
Department of Computer Science, St. Mary's College, Shillong, Meghalaya, India
e-mail: d.debnath@smcs.ac.in

S. Kr. Chettri
Department of Computer Applications, Assam Don Bosco University, Guwahati, India
e-mail: sarat.chettri@dbuniversity.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_11

revenue of blockchain-based enterprise applications is projected to be around $19.9 billion by 2025 [1]. Despite its great potential, blockchain technology faces certain issues, such as lower scalability, leakage of privacy, various types of attacks, large computing power requirements, and other technical challenges.

The rest of this article is organized as follows. Section 2 provides a quick overview of blockchain and its architecture. Section 3 describes the five typical blockchain application domains. Section 4 highlights some of the open issues and challenges in this area. Section 5 concludes the research work along with future directions.

## 2 Overview of Blockchain

### *2.1 Background*

Blockchain is an emerging field with tremendous scope for research. The blockchain tale dates back to early 1979, when Ralph C. Merkle, a computer scientist, presented the idea of immutably chaining blocks with a cryptographic hash function in his dissertation at Stanford University [2]. The concept is known as the Merkle hash tree. In 1991, the digital time-stamping of documents has been implemented [3] and later in 1992, the Merkle tree was incorporated to add more documents in a single block. Much of the relevance of blockchain technology is accredited to Satoshi Nakamoto, who designed and implemented the first cryptocurrency known as Bitcoin in 2008. A Bitcoin serves as a decentralized public ledger maintained by anonymous consensus over the network. He used blockchain technology and the hashlike method to time-stamp blocks without requiring them to be signed by a third trusted party [4]. The evolution of blockchain can be staggered into four phases based on their applications:

**Blockchain 1.0 (2009–2013)**: In this phase, blockchain technology was primarily used as cryptocurrencies for digital payments, currency transfer, and remittance. Bitcoin [5] being the most prominent example in this segment.

**Blockchain 2.0 (2013–2015)**: The key innovation here is ‘smart contract’ which is an agreement between two people or parties. A smart contract is an autonomous computer program designed to automatically facilitate, verify, and enforcement of digital contracts without any central authorities or third party. The term ‘smart contract’ was coined by a cryptographer and legal scholar Nick Szabo in 1996. Smart contracts are applied in numerous fields such as financial services and management, health care, property ownership, voting, insurance, and the Internet of Things (IoT), among many others. A well-known example of a platform that runs smart contracts is Ethereum [6], proposed by Vitalik Buterin in 2013 and officially launched in 2015.

**Blockchain 3.0 (2015–2018)**: DApps abbreviated for Decentralized Applications is the focus here. DApps utilizes decentralized storage and decentralized communication. A traditional app’s backend code runs on a centralized server, whereas DApps backend code runs on decentralized peer-to-peer networks. A DApp can have its frontend hosted on decentralized storages, unlike traditional apps. Ethereum was the

first blockchain that offered DApp development, followed by NEO, EOS, Stratis, Lisk, and many others [6].

**Blockchain 4.0: (2018 onwards)**: The fourth-generation blockchain is an emerging technology and using varied techniques such as AI, machine learning (ML) promises to deliver Industry 4.0-based applications [7]. Blockchain 4.0 ecosystems offer an Integrated Development Framework (IDE) and cross-blockchain compatibility meaning most DApps will run on multiple 4.0-level blockchains.

## *2.2 Blockchain Transaction*

The primary objective of blockchain technology is to build trust, transparency, traceability, and data integrity. A blockchain is a decentralized, time-stamped collection of immutable records called blocks distributed over peer-to-peer networks. Each block consists of a series of transactions that are secured and chained to each other using cryptographic hash functions. Blockchain technology, also known as Digital Ledger Technology [8], has three important properties: (1) Trustless: users in the blockchain system do not need to know or trust each other and there is no trusted central authority such as the banking system (2) Permissionless: the system allows any node to freely join or leave the network at will and (3) Censorship Resistant: the system does not ban any verified transaction in a blockchain network.

Figure 1 depicts the step-by-step process of a blockchain-based transaction involving user A making a transaction with another user B. The transaction is first broadcasted to the blockchain network. Each transaction is verified by the miner node(s). The miner creates a new block and adds it to the existing blockchain. The current blockchain is then distributed over the network so that the other nodes can update their ledger.

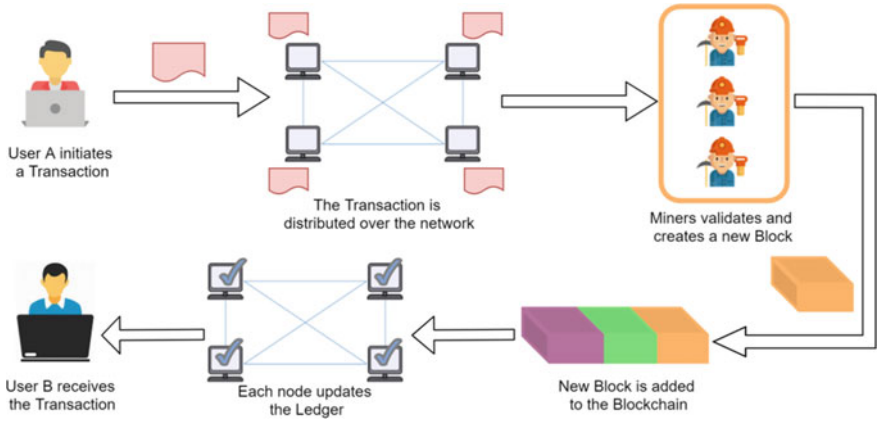



**Fig. 1** A stepwise process of blockchain-based transaction

## 2.3 Digital Signature

Digital signatures are an effective means of validating the authenticity and integrity of digital transactions in the blockchain environment. Digital signatures are implemented using public key cryptography based on the concept of public and private keys. Here each user has a pair of public and private keys. The public key is shared, and the private key is kept secret to the user. The private key is used to sign each transaction. The digitally signed content is then distributed over the network, which can be accessed and verified using the public key. Digital signatures provide a key feature called non-repudiation, where participants cannot deny that they have participated in the transaction. The most common algorithm used for digital signature blockchain is the elliptic curve digital signature algorithm (ECDSA). Other algorithms include SHA-256, SHA-3, Scrypt, X11, X13, and Ethash [9] among many others.

## 2.4 Blocks

A blockchain contains a sequence of blocks chained to a cryptographic hash function [10]. As shown in Fig. 2, each block is linked to its preceding block (termed as parent block). The first block that has no parent block is called the genesis block. Each block contains a block header, a transaction count, and several verified transactions. Block header includes metadata such as block version, block hash, parent block hash, target, Merkle root, time-stamp, and nonce. Table 1 summarizes the components of a block data structure.

Merkle Root [11] is the hash value of all validated block transactions. As shown in Fig. 2, all transactions are double-hashed to a hash value; then they combine

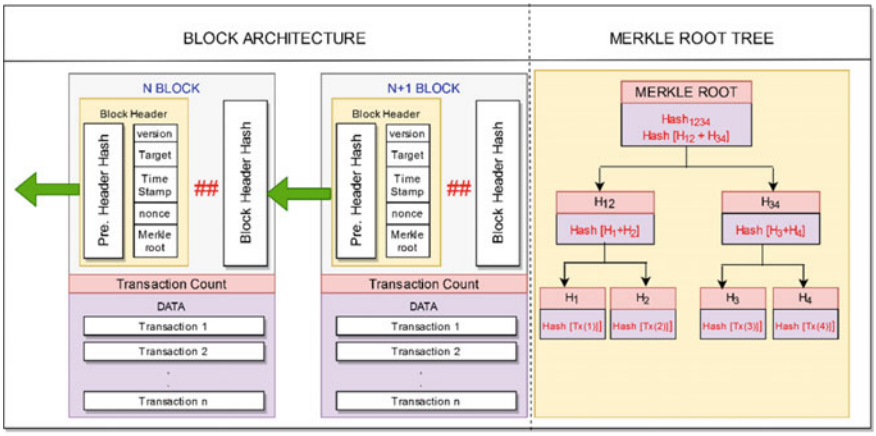



**Fig. 2** Block diagram of a block and Merkle tree

**Table 1** A summary of block data structure

| Data item | Data type | Byte | Description |
|---|---|---|---|
| Block hash | Char | 32 | Block hash is a 256-bit hash value of concatenated data fields of a block |
| Block version | Int32 | 4 | The block number, which governs the verification rules of a block set |
| Parent block hash | Char | 32 | A 256-bit block hash value of the parent |
| Time-stamp | Unit32 | 4 | Time-stamp in seconds since 1970-01-01T00:00UTC |
| Target | Unit32 | 4 | The difficulty target provided to the miner |
| Nonce | Unit32 | 4 | Nonce is an abbreviation for 'number only used once' is a value adjusted by miners to meet the target value |
| Merkle root | Char | 32 | A 256-bit hash value of all transactions of a block |
| Transaction counter | | 1–9 | Number of transactions |
| Transaction 1<br>Transaction 2<br>.<br>.<br>Transaction *n* | | Variable | Transaction data of the block |

pairwise and are entered into the parent hash function. The process is repeated until there remains a single hash value called the Merkle root or the Merkle digest. Merkle tree is also known as the binary hash tree and is used to optimally summarize and verify the integrity of large sets of block transactions.

## *2.5 Mining and Consensus Algorithms*

The process of adding transactions to the distributed blockchain ledger is called mining. It is a mechanism that allows the blockchain system to operate without a central authority. Miners are unique network nodes that solve a complicated mathematical problem to win the right to create a new block and to win a maximum reward in terms of Bitcoin [18]. Several algorithms decide which miner wins. These algorithms are referred to as consensus algorithms. The consensus is an automatic way of reaching an agreement between untrustworthy nodes to maintain consistency in the distributed ledger [19]. There are four major consensus algorithms such as (i) Proof of Work (PoW) [12], (ii) Proof of Stake (PoS) [13], (iii) Practical Byzantine Fault Tolerance (PBFT) [14], and (iv) Delegated Proof of Stake (DPoS) [15]. Some of the other approaches include Proof of Elapsed Time (PoET) [17], Proof of Space (PoSpace) [20], Proof of Importance (PoI), Measure of Trust (MoT), Minimum Block Hash (MBT), and Tindermint [16].

In PoW, miners need to solve a complex numerical puzzle to add a new block into the ledger. In PoW, miners need specialized hardware to solve the puzzle and win

**Table 2** A comparison table of different mining techniques

| Mining techniques | Energy consumption | Resource required | Adversary tolerance (%) | Consensus time | Scalability | Platforms |
|---|---|---|---|---|---|---|
| PoW [12] | High | Specialized hardware | ≤25 | High | Strong | Bitcoin, Litecoin |
| PoS [13] | Low | Wealth or stake | <51 | High | Strong | Ethereum, Peercoin |
| PBFT [14] | Very low | None | <51 | Low | Weak | Hyperledger fabric |
| DPoS [15] | Very low | Low wealth or stake | ≤33.3 | Medium | Strong | BitShares |
| Tindermint [16] | Very low | Permissioned | ≤33.3 | Low | Medium | Tindermint |
| PoET [17] | Medium | Low | ≤25 | Medium | High | Sawtooth, fabric |

rewards. PoS is a consensus algorithm where the miners (known as validators) lock-up or stake their crypto-coins as collateral to win the right to verify the transaction. PBFT is derived from Byzantine General's problem. According to PBFT, some amount of fault or wrong doing can be tolerated without affecting the integrity of the network. DPoS was developed by Daniel Larimer, is a variation of the Proof of Stack algorithm, and allows blockchain to change network parameters such as fee structure, block intervals, transaction sizes on the fly by the delegates who vote for such changes. A comparison of some of the popular mining techniques is shown in Table 2.

## 3 Applications of Blockchain

In diverse applications, blockchain technology is used. In this section, the five typical blockchain technology application domains were summarized. Most of the blockchain-based application areas are still in their infancy because most are at the conceptual level and need to do a lot for future blockchain systems.

### *3.1 Biomedical Domain*

There are various issues in areas under the biomedical domain [21] being addressed using blockchain technology. Some of the popular blockchain technologies used include Ethereum, Bitcoin, NEM, and Hyperledger Fabric among others. In the biomedical domain, blockchain technology is mainly used to maintain data integrity, enable distributed access control, data lineage, and non-repudiation. The data refers

to medical records, personal health records, personal biosensor data, clinical trial data, medical insurance, etc. Mostly, blockchain is used to address the issues of data security and privacy prevalent in the distributed environment. Other uses include the assurance of non-repudiation of medical acts or transactions and to build a reliable and transparent ecosystem in the healthcare sector. The integration of blockchain with the Internet of Things (IoT) in the biomedical domain solves the issues of reliability and data privacy and secure information related to the IoT paradigm. Likewise, the amalgamation of big data, machine learning, AI, blockchain, and edge computing is becoming an important component of the innovations in the biomedical domain.

### *3.2 Banking and Finance*

Blockchain became one of the banking sector's most popular technologies, as security is of utmost importance for the banking and finance domain [22]. In addition to security, blockchain offers operational advantages to banks, such as transparency, privacy, immutability, reduced costs, reduced human error, and quicker transactions. With different forms of payment such as stablecoin, tokenized fiat, and cryptocurrency, blockchain can reduce the remittance time with faster settlement times in domestic retail or cross-border payments. The blockchain reduces the risk of fraud and allows secured networks and distributed processes to generate more effective financing structures. With streamlined digital data verification and authentication, property and insurance claims can be revamped along with automated claims processing using smart contracts. It is possible to eliminate the necessity for a trusted third party when buying and selling stocks using blockchain technology, making it less expensive and faster transactions. Blockchain products challenge a broad range of traditional banking and financial products, but with blockchain technology, this sector is making a new paradigm shift in its services and products.

### *3.3 Government and Public Sectors*

Blockchain technology has a tremendous potential to enable smart governance [23]. A blockchain-based government model can provide seamless services to its citizens with trust and accountability because of the key features of blockchain: auditability and persistency. Some of these services include registration of assets (e.g., land registry, property ownership), systems for income taxation, patent management, identity management, and so on. The government can leverage blockchain technology to provide cybersecurity, process automation, secure storage of government's and citizens' data, cost reduction in accountability management, transparency in voting systems, protection of sensitive data in cyberspace, and corruption reduction, among others. Besides, blockchain has potential applications in energy conservation (green energy) and education (smart learning). Furthermore, with the introduction of smart

contracts, the government bureaucratic operations that are generally complicated can be executed in a streamlined manner.

### *3.4 Supply Chain*

A supply chain is a network that brings together individual entities, organizations, businesses, resources, and technologies to produce a product or service. Industry 4.0 tends to have a positive impact on supply chain management, supported by disruptive technologies such as big data analytics, machine learning, AI, 3D printing, cloud computing, and robotics, etc. Among these, as a distributed digital ledger, blockchain explores new ways to provide solutions [24] to various issues being faced by the supply chain industry in terms of transparency, traceability, and security. Some of the apparent benefits of using blockchain technology in supply chain management include real-time product tracking throughout its life cycle, a more transparent and traceable supply chain, improving product and service licensing, reducing the risk of fraud and product duplication, preventing tampering and building trust among suppliers, manufacturers, and customers, etc. However, there are certain challenges involved in adopting blockchains in supply chain management, such as lack of standards and protocols, legal issues, privacy issues, and error intolerance. A significant amount of collaborative work must be done by industry and academia for future research directions to understand and provide practical applications of blockchain and its performance measurement in the supply chain industry.

### *3.5 Internet of Things*

The Internet of Things (IoT) is applied in several diverse fields practically in all areas of the daily life of individuals, organizations, and society as a whole. Some of the important IoT application areas are in the fields of healthcare and medicine, agriculture and irrigation, automotive industry, manufacturing industry (e.g., Industrial IoT), public transportation, smart homes, etc. Using blockchain, various limitations of traditional IoT applications can be addressed, such as secure data transfer between IoT devices and applications, protection of sensitive information when delivered via the IoT network, ensuring the integrity of IoT data without a third party, optimizing the computational power of IoT devices, reducing the operating costs of IoT applications, etc. In reality, blockchain provides IoT devices, applications, and platforms with a scalable and decentralized environment. Moreover, blockchain-based IoT systems allow companies to run smart applications and execute multiple legal procedures between business partners through smart contracts. However, IoT systems suffer from problems of scalability, data security, and centralization and thus the convergence of blockchain and artificial intelligence for IoT applications

[25] together with cloud and fog computing could lead to the development of a secure, scalable, decentralized, and intelligent IoT ecosystem.

## 4 Open Issues and Challenges

Being an emerging technology, blockchain faces various issues and challenges. The three typical parameters are summarized in the sections: scalability in Sect. 4.1, attacks on blockchain in Sect. 4.2, and privacy breach in Sect. 4.3.

### *4.1 Scalability*

Scalability is one of the key limitations of blockchain technology. In fact, there is a key challenge in balancing the trade-off between the three key aspects of blockchain, commonly known as the Blockchain Trilemma, namely security, decentralization, and scalability. Balancing the trade-offs between these three aspects or to achieve all the three is important for the development of future blockchain systems. The problem of scalability arises due to the increasing number of transactions day by day, where there is a constraint in the network's throughput, i.e., the number of transactions that can be processed per second, for example, Bitcoin blockchain can only process 7 tx/s (transactions per second), Ethereum [26] processes 15 transactions per second while Visa and Paypal up to 1700 tx/s and 193 tx/s, respectively. Besides, with the limited block size (e.g., 1 MB in Bitcoin [4]) and the high latency or long confirmation time, i.e., the time needed to create a new block and its inclusion in the blockchain (e.g., around 10 min for Bitcoin's network and around 20 s for Ethereum's network), millions of transactions cannot be processed in real-time. Besides, throughput and latency, other factors that have an impact on scalability are storage and transaction fees paid to the miner. Currently, the blockchain size of Bitcoin (BTC) has exceeded 300 GB while Ethereum (ETH) and Litecoin (LTC) have exceeded 200 GB, and 28 GB of storage, respectively [27]. Several solutions have been proposed to address the scalability issue, such as network sharding [28, 29], consensus construction [30], lightning network [31], and directed acyclic graph [32].

Sharding is applied to blockchain to randomly break the network into small zones or shards with some teams of nodes assigned the responsibility. Thus, each node does not need to maintain the entire ledger to execute processes and validate transactions; instead, a part of the ledger is assigned to the nodes to be maintained to operate. This results in increased throughput as validation of transactions happens in parallel rather than in a linear fashion. So, scalability can be achieved with some compromise in security as hackers find it comparatively easier to take over a single shard, also known as a 1% attack. Various sharding protocols to solve the scalability issue in blockchain exist in the literature such as RSCoin [33], Chainspace [34], Elastico [29], OmniLedger [35], etc. A comparison of state-of-the-art sharding protocols

based on some parameters like safety aspects, consensus, protocol settings, and their performance can be found in the works by G. Wang et al. [28].

Such self-governing individual shards are asynchronized, so consensus among them becomes difficult to achieve. Consensus protocols are designed to make them reach an agreement on a shared state or data in a distributed environment. Broadly consensus algorithms can be classified into three broad categories [36]: (1) Incentivised consensus algorithms (Proof of Stake (PoS), Proof of Work (PoW), etc.), (2) Non-incentivised consensus algorithms that are mainly used in private blockchain systems for non-crypto-currency applications, and (3) Hybrid consensus algorithms (Proof of Research, Proof of Stake Velocity, Proof of Burn etc.). Designing and deploying a consensus protocol is another challenging task [37] because a lot of factors need to be considered such as network latency, safety, fault tolerance, transaction rate, efficient handling of corrupt inputs, and so on.

The change of the linear-chain structure of the blockchain with the DAG-based structure is another solution proposed in the literature to achieve greater scalability in the blockchain, where, through the directed acyclic graph, blockchain networks are extended and no mining occurs, it improves scalability with reduced transaction costs.

### *4.2 Attacks on Blockchain*

With the advent of blockchain-based applications beyond Bitcoin, it is natural to have a better understanding of different types of attacks on Blockchain technology, so that a strong foundation can be established to reduce attack opportunities pursuing better security. Private Blockchains are less susceptible to adverse attacks as compared to public Blockchains due to restricted access to system resources and stronger trust models. The attacks on the public blockchain can be categorized into three broad categories [38]: (1) Attacks based on the design constructs like the mathematical techniques used to build the Blockchain structure such as forks [39], orphaned blocks [40], etc. (2) Attacks based on the applications applying Blockchain technology like cryptojacking, double-spending [41], smart contract DoS, wallet theft, etc., (3) Attacks based on peer-to-peer architecture such as selfish mining [42], consensus delay, 51% attack [43], DDoS attack [44], eclipse attack, Domain Name System (DNS) attack and so on. The taxonomy of the attacks on public Blockchain is shown in Fig. 3.

Sometimes, to achieve better scalability and to reduce the probability of orphaned blocks prevalent in decentralized mining networks, centralization of the blockchain network is done which in turn makes the network more vulnerable to attacks like double-spending and majority attacks or 51% attack where a group of miners could control more than 51% of network's mining hash rate or computing power. The 51% attack is one of the biggest threats to any Bitcoin-like currency. For example, in the year 2018, 51% of the network's mining hash rate in Bitcoin Gold (BTG) was acquired by malicious miners, stealing cryptocurrency costs of $18 million [45]

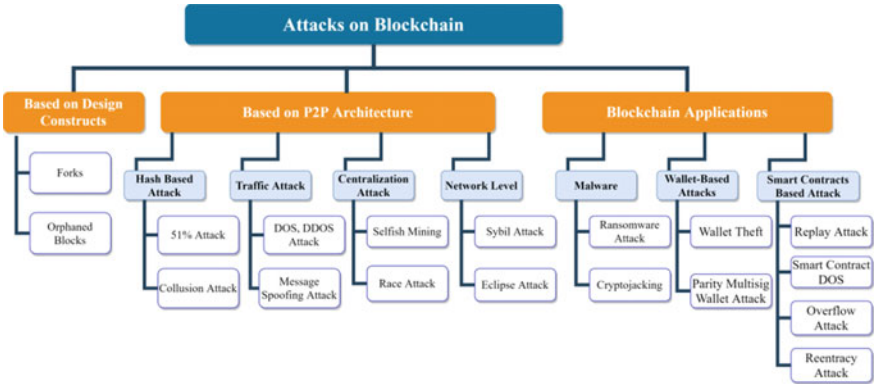



**Fig. 3** Taxonomy of attacks on blockchain

followed by $72,000 in the year 2020. Another issue is the fork (hard fork or soft fork) problem, where an attack is based on the design construct of the blockchain technology. The fork problem exists because of the issues in the compatibility of the consensus between the old node and new node verification requirement in the blockchain network. These blockchain forks may cause a delay in consensus in the network leading to the double-spending attack. In recent studies, it is found that even nodes having computation power of less than 51% are still dangerous as there is a possibility of an attack known as selfish mining. The strategy applied in selfish mining by the miners is that they hide their mined blocks without broadcasting them to the public unless some requirements are satisfied. In this way, selfish miners keep mining their private chain without any competitive miners while honest or rational miners waste their resources mining the useless public chain. Attackers may launch eclipse attacks and DNS hijacks where the users and miners are kept isolated from the actual network. The attackers poison the DNS cache and modify the data and when any user makes any DNS query, they are led to some counterfeit networks. Another most common attack in the blockchain is the distributed denial-of-service (DDoS) attack. It is an attack made on blockchain-based applications such as Bitcoin and Ethereum. Even the 51% attack can cause the denial-of-service (DoS).

Another most challenging attack in the permissionless public blockchain is the Sybil attack [46], where attackers can carry out a 51% attack. In the Sybil attack, the main idea followed is that attackers create fake identities or Sybil identities in a blockchain network and use them to gain an unreasonably large influence over peers. Thus, they can take control of the hash rate of the network and could prevent confirmation of other transactions. The smart contracts are completely automated contracts, but smart contract-based attacks are due to the presence of bugs, and one of the biggest exploitation in the history of cryptocurrencies is the 'THE DAO' (distributed automated organizations) hack in the year 2016 where an attacker could pull out almost $70 million out of the crowdfunding project.

### *4.3 Privacy Breach*

As users perform transactions in anonymity without revealing their real identity, the blockchain technology retains a high degree of user's privacy. However, Kosba et al. [47] have stated that even though the transactions use pseudonymous addresses, user's privacy in the Bitcoin blockchain cannot be guaranteed. And once the user's true identity is revealed, his transactions are seriously threatened of getting leaked to the public. Various methods for improving the privacy of future blockchain systems are suggested by the researchers. These techniques mainly use anonymization methods such as MixCoin, CoinJoin, CoinShuffle, etc., or cryptographic techniques such as Homomorphic Encryption (HE), Attribute-based Encryption, Non-Interactive Zero-Knowledge Proof (NIZK), etc.

The anonymization techniques such as Mixcoin [48] and CoinShuffle [49] enable users to make anonymous payments in Bitcoin or similar cryptocurrencies. However, the Bitcoin is pseudonymous rather than anonymous. In the proposed methods, users mix their coins with others to preserve privacy. These mixing strategies are either fully centralized where all users trust a mixer or a decentralized one without any requirement of trusted third parties like as proposed in CoinJoin [50] method developed for Bitcoin by Greg Maxwell. These methods are effective in preventing the identity of users from being disclosed and linked to, however, centralized services have a risk of privacy breach of the users and require a third party to be trusted.

In the cryptographic approach, data are stored in encrypted form over the blockchain without any substantial changes in its properties. It addresses the issue of privacy breach prevalent in public blockchains. For example, in Ethereum smart contracts, data are stored in the blockchain in encrypted form using homomorphic encryption techniques, to maintain the user's privacy. Secure multi-party communication (SMP) in a decentralized computation platform such as Enigma [51] is also proposed to provide a protection mechanism of personal data without any need to trust the third party to carry out joint computing by multiple parties in the blockchain network. The methods, however, suffer from a drawback as they only support simple operations such as additions and subtraction and suffer from lower computational efficiency for complex operations.

## 5 Conclusion

The traditional industry has immensely benefited from the six key features of blockchain: decentralization, transparency, anonymity, immutability, autonomy, and auditability. With an increasing interest in the application of blockchain technology at the industrial level and subsequently in academic research, the issues and challenges in blockchain are also increasing. This article reviews the current research and focuses primarily on the blockchain-based applications in various fields, ranging from finance to public services, from supply chain management to healthcare, highlighting open

issues and challenges with potential research opportunities for future blockchain-based systems and applications. It can be concluded that a deeper understanding of the key features of blockchain plays a vital role in addressing related issues such as data security and privacy, scalability, blockchain attacks, etc. and helps to develop technological innovation for future blockchain systems.

## References


1. Zhang R, Xue R, Liu L (2019) Security and privacy on blockchain. ACM Comput Surv 52(3):1–34
2. Sherman AT, Javani F, Zhang H, Golaszewski E (2019) On the origins and variations of blockchain technologies. IEEE Secur Privacy 17(1):72–77
3. Haber S, Scott Stornetta W (1991) How to time-stamp a digital document. J Cryptogr 537(3):437–455
4. Nakamoto S (2009) Bitcoin: a peer-to-peer electronic cash system. Cryptogr Mail List, pp 1–9. https//metzdowd.com
5. Xu M, Chen X, Kou G (2019) A systematic review of blockchain. Financ Innov 5(1):5–27
6. Wang S, Ouyang L, Yuan Y, Ni X, Han X, Wang FY (2019) Blockchain-enabled smart contracts: architecture, applications, and future trends. IEEE Trans Syst Man Cybern Syst 49(11), 2266–2277
7. Alladi T, Chamola V, Parizi RM, Choo KKR (2019) Blockchain applications for Industry 4.0 and industrial IoT: a review. IEEE Access 7:176935–176951
8. Wang X, Wang L (2019) Parallel and distributed computing, applications and technologies. In: 19th international conference, PDCAT 2018, 20–22, Jeju Island, South Korea. Springer, Singapore
9. Zheng Z, Xie S, Dai HN, Chen X, Wang H (2018) A lightweight hash-based blockchain architecture for industrial IoT. Int J Web Grid Serv 14(4):1–17
10. Bakhtiari S, Pieprzyk J, Safavi-Naini R (1995) Cryptographic hash functions: a survey. Technical Report, pp 95–109
11. Zhang J, Zhong S, Wang T, Chao HC, Wang J (2020) Blockchain-based systems and applications: a survey. J Internet Technol 21(1):1–14
12. Dey S (2019) Securing majority-attack in blockchain using machine learning and algorithmic game theory: a proof of work. In: 2018 10th computer science and electronic engineering (CEEC) CEEC 2018—proceedings, pp 7–10
13. Li W, Andreina S, Bohli JM, Karame G (2017) Securing proof-of-stake blockchain protocols. Lecture notes in computer science (including Subseries Lecture notes in artificial intelligence (LNAI) and lecture notes in bioinformatics), vol 10436. LNCS, pp 297–315. https://doi.org/10.1007/978-3-319-67816-0_17
14. Abraham I, Gueta G, Malkhi D, Alvisi L, Kotla R, Martin JP (2017) Revisiting fast practical byzantine fault tolerance. Preprint arXiv:1712.01367v1 [cs.CR], no. i, pp 1–13 [online]. Available at: http://arxiv.org/abs/1712.01367
15. Yang F, Zhou W, Wu Q, Long R, Xiong NN, Zhou M (2019) Delegated proof of stake with downgrade: a secure and efficient blockchain consensus algorithm with downgrade mechanism. IEEE Access 7:118541–118555
16. Amoussou-Guenou Y, Del Pozzo A, Potop-Butucaru M, Tucci-Piergiovanni S (2019) Correctness of tendermint-core blockchains. In: Leibniz international proceedings in informatics (LIPIcs), vol 125, pp 1–30
17. Ali Syed T, Alzahrani A, Jan S, Siddiqui MS, Nadeem A, Alghamdi T (2019) A comparative analysis of blockchain architecture and its applications: problems and recommendations. IEEE Access 7:176838–176869

18. Zheng X, Zhu Y, Si X (2019) A survey on challenges and progresses in blockchain technologies: a performance and security perspective. Appl Sci 9(22):1–24
19. Wan S, Li M, Liu G, Wang C (2019) Recent advances in consensus protocols for blockchain: a survey. Wirel Netw 26(8):5579–5593
20. Park S, Kwon A, Fuchsbauer G, Gaži P, Alwen J, Pietrzak K (2018) SpaceMint: a cryptocurrency based on proofs of space. Lecture notes in computer science (including Subseries Lecture notes in artificial intelligence (LNAI) and lecture notes in bioinformatics), vol 10957. LNCS, pp 480–499. https://doi.org/10.1007/978-3-662-58387-6_26
21. Drosatos G, Kaldoudi E (2019) Blockchain applications in the biomedical domain: a scoping review. Comput Struct Biotechnol J 17:229–240
22. Osmani M, El-Haddadeh R, Hindi N, Janssen M, Weerakkody V (2020) Blockchain for next generation services in banking and finance: cost, benefit, risk and opportunity analysis. J Enterp Inf Manag
23. Alketbi A, Nasir Q, Talib MA (2018) Blockchain for government services—use cases, security benefits and challenges. In: 15th learning and technology conference (L&T), pp 112–119
24. Saberi S, Kouhizadeh M, Sarkis J, Shen L (2019) Blockchain technology and its relationships to sustainable supply chain management. Int J Prod Res 57(7):2117–2135
25. Singh SK, Rathore S, Park JH (2020) BlockIoTIntelligence: a blockchain-enabled intelligent IoT architecture with artificial intelligence. Future Gener Comput Syst 110:721–743
26. Wood G (2014) Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj. Yellow Paper, pp 1–32
27. Bitinfocharts.com. Cryptocurrency statistics. https://bitinfocharts.com/. Accessed 15 Sept 2020
28. Wang G, Shi ZJ, Nixon M, Han S (2019) Sok: sharding on blockchain. In: 1st ACM conference on advances in financial technologies, pp 41–61
29. Luu L, Narayanan V, Zheng C, Baweja K, Gilbert S, Saxena P (2016) A secure sharding protocol for open blockchains. In: ACM conference on computer and communications security, pp 17–30
30. Du M, Ma X, Zhang Z, Wang X, Chen Q (2017) A review on consensus algorithm of blockchain. In: 2017 IEEE international conference on systems, man, and cybernetics, SMC 2017, Jan 2017, pp 2567–2572
31. Poon J, Dryja T (2016) The bitcoin lightning network: scalable off-chain instant payments. Percept Psychophys 18(3):205–208
32. Pervez H, Muneeb M, Irfan MU, Ul Haq I (2019) A comparative analysis of DAG-based blockchain architectures. In: ICOSST 2018—2018 international conference on open source systems and technologies—proceedings, pp 27–34
33. Danezis G, Meiklejohn S (2015) Centrally banked cryptocurrencies. arXiv Preprint arXiv:1505.06895, pp 1–15
34. Al-Bassam M, Sonnino A, Bano S, Hrycyszyn D, Danezis G (2017) Chainspace: a sharded smart contracts platform. Preprint arXiv:1708.03778v1, pp 1–16
35. Kokoris-Kogias E, Jovanovic P, Gasser L, Gailly N, Syta E, Ford B (2018) OmniLedger: a secure, scale-out, decentralized ledger via sharding. In: IEEE symposium on security and privacy, pp 583–598
36. Ferdous MS, Chowdhury MJM, Hoque MA, Colman A: Blockchain consensus algorithms: a survey. Preprint arXiv:2001.07091v2 [cs.Dc], pp 1–39 [online]. Available at: http://arxiv.org/abs/2001.07091
37. Baliga A Understanding blockchain consensus models [online]. Available at: https://www.persistent.com/wp-content/uploads/2017/04/WP-Understanding-Blockchain-Consensus-Models.pdf
38. Saad M et al (2019) Exploring the attack surface of blockchain: a systematic overview. Preprint arXiv:1904.03487v1, pp 1–30 [online]. Available at: http://arxiv.org/abs/1904.03487
39. Eyal I (2015) The miner's dilemma. In: IEEE symposium on security and privacy, pp 89–103. https://doi.org/10.1109/SP.2015.13
40. Decker C, Wattenhofer R (2013) Information propagation in the bitcoin network information propagation in the bitcoin network. In: IEEE international conference on peer-to-peer computing, vol 13, pp 1–10

41. Dilhani I, De Zoysa TN (2017) Transaction verification model over double spending for peer-to-peer digital currency transactions based on blockchain architecture. Int J Comput Appl 163(5):24–31
42. Leelavimolsilp T, Tran-Thanh L, Stein S: On the preliminary investigation of selfish mining strategy with multiple selfish miners. Preprint arXiv:1802.02218v1 [cs.MA], pp 1–20 [online]. Available at: http://arxiv.org/abs/1802.02218
43. Bastiaan M: Preventing the 51%-attack: a stochastic analysis of two phase proof of work in bitcoin. Available at: http://referaat.cs.utwente.nl/conference/22/paper/7473/preventingthe-51-attack-astochastic-analysis-of-two-phase-proof-of-work-in-bitcoin.pdf. Accessed 20 Sept 2020
44. Saad M, Thai T, Mohaisen A (2018) POSTER: deterring DDoS attacks on blockchain-based cryptocurrencies through mempool optimization. In: ASIACCS 2018—Proceedings of the 2018 ACM Asia conference on computer and communications security, pp 809–811
45. Roberts J (2018) Bitcoin gold suffers rare '51% attack.' Fortune. http://fortune.com/2018/05/29/bitcoin-gold-hack/. Accessed 26 Sept 2020
46. Otte P, de Vos M, Pouwelse J (2020) TrustChain: a Sybil-resistant scalable blockchain. Future Gener Comput Syst 107:770–780
47. Kosba A, Miller A, Shi E, Wen Z, Papamanthou C (2016) Hawk: the blockchain model of cryptography and privacy-preserving smart contracts. In: Proceedings—2016 IEEE symposium on security and privacy, SP 2016, pp 839–858
48. Bonneau J, Narayanan A, Miller A, Clark J, Kroll JA, Felten EW (2014) Mixcoin: anonymity for bitcoin with accountable mixes. In: Proceedings of the IEEE symposium on security and privacy (SP), San Jose, CA, USA, vol 8437, pp 486–504
49. Ruffing T, Moreno-Sanchez P, Kate A (2013) CoinShuffle: practical decentralized coin mixing for bitcoin. In: Proceedings—European symposium on research in computer security, vol 8713. LNCS, PART 2, pp 345–364. https://doi.org/10.1007/978-3-319-11212-1_20
50. Maxwell G (2020) CoinJoin: bitcoin privacy for the real world. In: Post on bitcoin forum 2013. bitcointalk.org [online]. Accessed 20 Sept 2020
51. Zyskind G, Nathan O, Pentland A (2015) Enigma: decentralized computation platform with guaranteed privacy. Preprint arXiv:1506.03471v1 [cs.CR], pp 1–14. https://doi.org/10.7551/mitpress/11636.003.0018

# A Study of Mobile Ad hoc Network and Its Performance Optimization Algorithm


**Vishal Polara and Jagdish M. Rathod**

**Abstract** MANET is basically used for quick transmission of data without creating infrastructure-based network. In order to do transmission in mobile ad hoc network, it is required to find out efficient path between source and destination. MANET is developed using different type of topologies which frequently get changed on the movement of node that affects network performance result in selection of inappropriate path. So to find out best path between nodes, routing protocol is required. There are number of routing protocols available based on the requirement of network it get selected and implemented. This paper provides the information of various routing protocol and optimization algorithm developed by different researcher which is applied to improve performance of MANET.

**Keywords** Mobile ad hoc network · Topology · Protocol · Ad hoc on-demand vector · Dynamic source routing · Zone routing protocol · Particle swarm optimization

## 1 Introduction

### *1.1 MANET*

A wireless ad hoc network is a group of mobile nodes without having preplanned infrastructure, which form a temporary network. Each and every node is communicating with each other using radio or infrared communication technology. Laptop and mobile or let us say personal digital assistants which can communicate directly are the example of ad hoc network. Node is normally mobile in ad hoc networks.

Ad hoc network contains few stationary nodes for example access point to the Internet, and there are also few semi-mobile nodes available for deploying relay point

V. Polara (✉) · J. M. Rathod
BVM Engineering College, Anand, Gujarat, India

J. M. Rathod
e-mail: jmrathod@bvmengineering.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_12

in the area where there is a need of relay point temporarily. The outsider nodes are normally not within the transmitter range of each other. The center nodes are most useful to transfer the packet between two nodes. It acts as a router so that transmission between two nodes can establish successfully.

In an ad hoc network, there is no central administrator available for handling the packet. Network does not get collapsed because of frailer of one node or the node goes outside the network range. Node can enter and leave the network to satisfy their requirement. Nodes are having limited transmission capability so if two nodes want to establish connection they required to use multihop approach in which other nodes act as an intermediate node. Every node must be willingly transmitting the packets for other nodes [1]. So every node acts as a router and host simultaneously. A node can be defined as an abstraction of router and a number of mobile nodes which are acting as a hosts.

A router is defined as a node which has the capability to store the entry of all the nodes to which packet is transmitted. It also provides optimum path to reach one node to another node. A node is nothing but a host which contain an IP address.

In an ad hoc network, topologies are changing rapidly and there are also various possibilities of malfunctioning of nodes but ad hoc network is capable to handle all possible problems as mentioned. It fixed everything with the help of network configuration features.

If node wants to communicate with each other, they require routing protocol. Routing protocol comes with two main functions: First function is the selection of path for numbers of source and destination node available for transmission of packet at right destination. The second function is to decide and maintain routing table so that with less number of network problem or link failure transmission could be achieved.

## *1.2 Characteristics of MANET*

MANET can be described by following characteristics [4].

**No Need of Infrastructure**: Mobile ad hoc network is created without having central administrator as it does not depend on any pre-established infrastructure. Here, all the nodes communicate with each other using peer-to-peer approach and with the help of their own router to generate data. Management functionality of network is distributed along various nodes. It also helps to manage fault detection. There is no discrimination between endpoints and switches in case of MANET.

**Multiple hop routing**: In MANET, there is no fixed router available for routing information between nodes. Here, node itself works as a router and forward information between various mobile hosts. Usually two types of routing take place like single hope and multihope based on various layer attributes and routing protocols. In single-hop routing implementation mechanism is very simple with lost cost and less functionality. In multihope mechanism, data packets use direct wireless transmission

protocol and range to forward packet from source to destination by taking more than one node as an intermediate nodes.

**Dynamic topology**: In MANET, nodes are moving randomly so topology of node is not fixed it is usually dynamic in nature. Normally in multihope network implementation topologies are changing rapidly and unpredictably which will result in frequent route change and partition of network with loss of packet. MANET can able to work in various propagation conditions and dynamic type of traffic as well as various types of mobility patterns. Node is creating different types of network as they are continuously moving on the fly mode. User is required to operate not only from mobile ad hoc network but also required to have public network which is normally fixed like Internet.

**Variation of Node and link capabilities**: All nodes are having one or more radio interfaces with different transmission and receiving capabilities as well as operating with different frequency bands. The rate of bit error of wireless connection is usually more profound in mobile ad hoc network. One direct path between two nodes is shared by the sessions. Nodes normally communicate through channel and channels are noisy or equipped with less bandwidth than a wired network. In certain case, node uses multiple wireless link for transmission.

**Low-weight terminals**: Mostly mobile nodes in mobile ad hoc network have low CPU processing speed with limited storage capability and also are compact in size. So these types of devices need an optimization algorithm to implement and manage computing and communication functions.

**Dynamic topology**: Nodes are moving arbitrarily across the network using random way point mobility model.

**Limited Bandwidth**: Node are available with limited bandwidth in MANET compared to wired network because of that less amount of throughput can be achieved.

**Energy-constrained operation**: One of the most important design considerations is node life. Node is equipped with the limited life battery in MANET.

**Security**: ad hoc networks are more exposed to problems compared to wired network which increased the chances of security attacks like denial of service and eavesdropping which needs to be handled efficiently.

## *1.3 Challenges in Ad hoc Network*

**Routing**: Dynamic topology is used for routing in MANET. So routing of packet either by using any approach becomes more challenging compared to wire network. Mostly protocol which is used in network is reactive in nature [2].

**Routing Overhead**: In MANET, nodes are changing their location rapidly so the route which is generated through random movement creates a routing overhead.

**Interference**: when nodes are communicating with each other links are established and fail depending on the various characteristics of transmission which may result in overhearing of conversation by other nodes and which may also result in corruption of complete transmission.

**Security and Reliability**: ad hoc networks are less secure because of the problem mostly seen as to, e.g., depending on neighbor relaying packets. In MANET, links are having characteristics to introduce reliability problems, because of limitation of transmission range, medium is broadcast in nature, packet loss because of mobility and data loss on transmission [4].

**Asymmetric links**: Usually wired networks depend on the fixed symmetric links. But it is not with ad hoc networks as the nodes are randomly moving using dynamic topology and changing their position frequently within the network, e.g., when node A sends packets to node B it does not tell anything about reverse direction link.

**Internetworking**: MANET- and IP-based networks are usually communicating with each other. So the correlation of routing protocols between mobile devices is challenging for the harmonious mobility management.

**Dynamic Topology**: It has major problem of routing because topologies are dynamic in nature. In MANET, type of network routing table must reflect this change of topology and routing algorithms effectively, e.g., routing table updated after every 30 s in case of fixed network. Frequency of updating in ad hoc network is low compared to fixed network.

**Energy Consumption**: MANET nodes are operating usually on the batteries, so the power conservation becomes a crucial problem. Hence, power consumption must be optimized. Conservation of power-aware routing must take place [15].

## 2 Overview of routing protocol

Protocols are rules that govern the communication. Routing protocols are necessary in order to establish trustworthy communication. In mobile ad hoc network, there are basically three types of routing protocols proactive, reactive, and hybrid. Explanation of all categories of protocol is explained in detail.

### *2.1 Classification of Routing Protocol*

Figure 1 shows the classification of routing protocol.

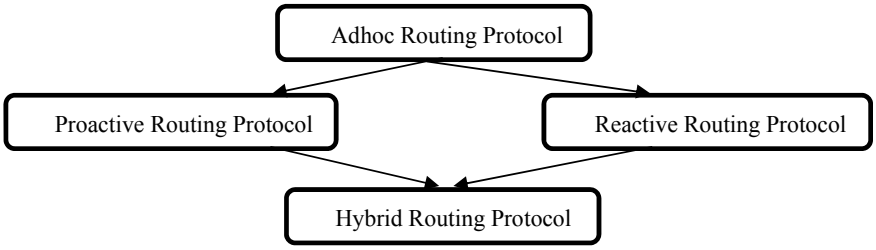



**Fig. 1** Ad hoc routing protocol

- ***Reactive routing protocols***

It is also famous as a demand routing protocol. In this type of routing protocol, route discovery is based on demand. Nodes discover route on demand basis. Source node checks available route in routing table if proper route is note available it will initiate the route discovery mechanism [10].

***Route discovery***: In this process, route is discovered by source when it is required. Source node checks buffer for the availability of route from source to destination if the route is not present it starts the process of discovering new route. Source node contains complete path to reach destination and it also contains the information of intermediate nodes.

***Route maintenance***: because of network having dynamic topology there are more number of link failure that occurs so there is a requirement of proper route maintenance. Reactive protocols come with the acknowledgement mechanism because of that maintenance of route is possible. Delay is added to the network because of implementation of route discovery mechanism. Each intermediate node also adds latency which is generated in the route discovery process [5]. Routing overhead is reduced in this type of protocol but the cost of delay in network gets increase. Therefore, this type of protocols is useful when network required low routing overhead. There are many famous reactive routing protocols available in MANET, for example, DSR, AODV, TORA, and LMR.

- **Proactive routing protocols**

It is also popular as a table-driven protocol. In this mechanism, each and every node maintains one routing table which contains information about topology of network even without the requirement of it. This feature is also useful for various types of traffic like data, substantial signaling traffic, and power consumption [3, 4]. It periodically updates the routing table at the time of change in network topology. It is not preferable to use for large network as they need to maintain entry of each and every node in the routing table [6]. There are various types of protocol and each protocol has a different number of routing tables.

There are many popular proactive routing protocols available like OLSR, DSDV, and WRP. Destination sequence routing protocol (DSDV) is a table-driven routing protocol based on the distributed Bellman–Ford algorithm. There are few improvement made in existing Bellman–Ford algorithm by making it free from loop in routing tables [1]. In this algorithm, each and every node maintains routing table which contains numbers of hope to reach the destination, next hop, and sequence number. This information is attached to each node. Destination sequence routing protocol has large overhead because of routing table. Enhanced version of DSDV protocol is wireless routing protocol (WRP). It maintains all the information related to routing in routing table because of the nature of proactive routing.

Mainly four types of tables are maintained in this protocol namely routing table, distance table, cost table of links, and retransmission list of message.

- ***Hybrid routing protocol***

Mostly protocols presented here are of either proactive type or reactive type. There is some difference in working of proactive and reactive routing protocol. Usually proactive routing protocol has less latency and large overhead while reactive routing protocols have more latency and less overhead. Hence, hybrid protocol is used to overcome the drawback of proactive and reactive routing protocols.

It combined features of both the routing protocol [3]. It uses the table maintenance mechanism of proactive protocol and the route discovery mechanism of reactive protocol so as to avoid latency and overhead problems in the network. It is also suitable for large networks where large numbers of nodes are present.

In this type of protocol, larger network is divided into numbers of zones where routing inside the zone is performed with the help of reactive routing protocol and outside the zone it is done using proactive routing protocol.

There are most popular hybrid routing protocols in MANET, e.g., ZRP and SHRP [7].

In the above section, discussion about routing protocol is carried out. When routing protocols are implemented in real-time environment, lots of issues are created like link failure, power failure, inappropriate execution of path, packet loss; so to overcome all this drawback, it requires to combine protocol with optimization algorithm so overall performance of network increases with respect to parameters like routing overhead, end-to-end delay deduction and also packet delivery ration can increase so in next section discussion about certain algorithm is given which can serve the above purpose.

## 3 Network performance optimization algorithm

Earlier discussion of MANET contains routing protocol supported by MANET, its advantages, and the problem associated with MANET network. Here, discussion

about various algorithms which can be used to enhance performance of ad hoc network is given.

## *3.1 Genetic algorithms (GA)*

It was developed by the Netherlands in 1975. It is working on a principal of natural selection and it is also a branch of a computational model. This method of optimization of network performance is the powerful one among the others.

Genetic algorithms are affected by humanism. Normally genetic algorithms work properly for optimization and are also called function optimizers. In this type of population solution, it is called chromosome, as it is initialized for algorithm.

Fitness is evaluated for each chromosome with the help of appropriate fitness function [8]. The best chromosome is chosen and crossover and mutation occur for better offspring. Genetic algorithm efficiently and conveniently works in the following case.

i. Search space is big, complex, and most probable not well known.
ii. There is no requirement of mathematical analysis.
iii. Knowledge of domain is scarce less to encode in order to narrow the searching area.
iv. It is used for complicated problems and the problem which are loosely coupled, it works with its own internal rules.
v. Searching method conventional approach fails.

## *3.2 Particle Swarm Optimization (PSO)*

It is probabilistic optimization method which is based on population, and it is proposed by Kennedy and Eberhart in 1995. Grouping behavior of birds and fish swimming inspired to develop this technique. In this technique, each member is represented by particles with independent speed and position. The best position is determined by the highest fitness value of particles [8]. It contain various step in algorithm to identify best value.

Following are the steps to decide best value:

i. Particles are initialized in search space
ii. Performance of each and every particles is evaluated.
iii. If the fitness value of particle is better than pbest value, then pbest value is set as a new value of particle.
iv. The position and velocity of particles updated.

### 3.3 Ant Colony Optimization (ACO)

This technique is inspired by ant feeding behavior, and it is also a part of metaheuristic technique. This approach was developed by Dorgio and Dicario in 1999. It has following three main function:

i. Ant Solution Construct: In this, non-natural ants move during adjacent states of predicament.
ii. Pheromone Update: After making clarification pheromone trails are restructured.
iii. Daemon actions: In this supplementary pheromone is applied to the superlative solution.

### 3.4 Artificial Bee Colony Optimization (ABC)

This algorithm is based on the working style of nature bees based on this numbers of algorithm available based on intelligence. This technique is based on foraging behaviors of bee group and it was proposed by Basturk and Karaboga. These algorithms divided into two categories: breeding behavior and mating behavior. In the ABC technique, there are groups of bees:

i. Spectator
ii. Utilize.

### 3.5 Bacterial Foraging Optimization Algorithm (BFOA)

It is a universal optimization algorithm inspired by feeding behavior of bacteria named Escherichia coli. Bacterial foraging optimization algorithm is affected by the chemotaxis of bacteria. These bacteria use concept of gradient of chemicals to obtain a direction to food. Strategy of information processing is achieved by a series of processes.

i. Chemotaxis: Shifting of cells along with the exterior one at a time.
ii. Reproduction: Sets of bacteria are chosen on the principal of best value to so it can be donates to the subsequent generation.
iii. Elimination and Dispersal: Cells are unnecessary and new illustrations are interleaved.

### 3.6 Binary Particle Swarm Optimization (BPSO)

In accumulation, there is a comprehensive description of PSO called BPSO. In binary particle swarm optimization, each particle has its binary value equal to 0 or 1. Value of

particle gets changed based on a movement from 0 to 1. In BPSO, velocity of particles is defined using the probability which will change its state to 1. It is useful in many applications like iterative prisoner's dilemma, optimal input subset of SVM, dual-band and dual polarization planar antenna design. If BPSO with PSO is compared then as the BPSO algorithm was used in the binary discrete search space, computational complexity of BPSO is reduced compared to PSO and also the accuracy of the calculation is reduced. BPSO has a finite state solution and can shorten the overall time required for calculation for particle convergence compared to PSO so it is a main advantage of BPSO over PSO. Qualities of binary PSO over particle swarm optimization are[14]:

i. The computational complexity get condensed
ii. The accuracy of calculation increased
iii. State resolution is restricted
iv. Easy development.

In the next section, discussion about one of the algorithms from above algorithm is explained in detail because from earlier studies it is proved that among all above algorithms PSO gives better performance. In earlier discussion of MANET, routing protocol is supported by MANET and its advantages and the problem are associated with MANET network. Here, discussion about various algorithms which can be used to enhance performance of ad hoc network is elaborated.

## 4 PSO Overview

It was developed by Kennedy and Eberhart in 1995 based on swarm behavior in nature, such as fish and bird schooling.

In place of using the crossover or mutation or pheromone, it uses real number randomness and global communication among the swarm particles. That makes this algorithm easier to implement as parameters does not require to encode and decode into binary strings compared to those in genetic algorithm where it uses a real number [12].

An objective function search space is obtained by adjusting the trajectories of individual agents, called particles, as the piecewise paths formed by positional vectors in a quasi-stochastic manner. A swarming particle movement consists of two major components:

- A stochastic component;
- A deterministic component.

Particles are attracted toward the current global best $g^*$ position and its own best location $x_i$ * in history. It has capacity to move randomly at the same time. If particle $i$ finds new best location than the existing one then it updates the available best location with existing one [13]. All $n$ particles have current best location at any time $t$ during

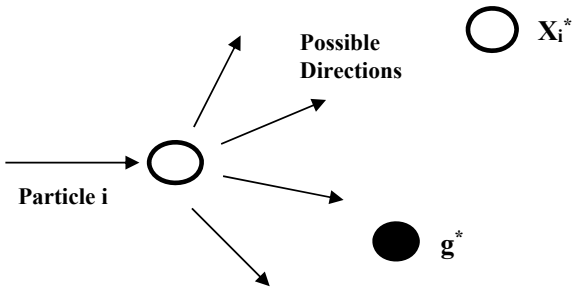



**Fig. 2** Notion of particle in PSO

iterations. This process of finding global best is repeated till the objective is achieved or for a specific number of iterations [9].

Particles movement is represented in Fig. 2, where $x_i$ *$(t)$ is the current best for particle $i$ and $g^* \approx \min\{f(x_i)\}$ for $(i = 1, 2, \ldots, n)$ is the current global best at $t$.

The following formulas are used to determine the new velocity and position vectors of the particle [11].

$$V_i^{t+1} = V_i^t + \alpha\varepsilon_1\left[g * -x_i^t\right] + \beta\varepsilon_2\left[x_i^{*(t)} - x_i^t\right] \tag{1}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{2}$$

where

- $x_i^t$ is the position vector for particle $i$ at instant $t$.
- $v_i^t$ is the velocity vector for particle $i$ at instant $t$.
- $\varepsilon_1$ and $\varepsilon_2$ are two random vectors, and each entry takes the values between 0 and 1.
- $\alpha$ and $\beta$ are the learning parameters or acceleration constants, which can typically be taken as, say, $\alpha \approx \beta \approx 2$.

The essential steps of the PSO can be summarized as the pseudocode below:

**Particle Swarm Optimization**

Objective function f (x), x = $(x_1 ... x_d)^T$

Initialize locations $x_i$ and velocity $v_i$ of n particles.

Find g* from min {$f(x_1) ... f(x_n)$} (at t=0)

**while (criterion)**

**for** loop over all d dimensions and all n particles

Create new velocity $v_i^{t+1}$

Find out new locations $x_i^{t+1} = x_i^t + v_i^{t+1}$

Apply objective functions at new locations $x_i^{t+1}$

Each particle $x_i$* current position is determine

**end for**

Obtain the current global best g*

Update t=t+1 (pseudo time or iteration counter)

**end while**

Output the final results $x_i$* and g*.

At the initial stage, distribution of all particles is relatively uniform because of that they can make sample over most of the regions, which is very important for multimodal type of problems. Velocity of particle can be taken as zero at the initial stage, that is,

$$v_i^{t=0} = 0.$$

## 5 Conclusion

This paper contains information about popular routing protocol used in MANET for communication. Each protocol functionality is provided and it is observed that most of the protocol face difficulty in finding optimal path to attain optimized performance in routing. Discussion includes certain algorithms which are widely used for improving network performance. The use of protocol is it depends on the requirement of network and the kind of parameter that would like to improve like link failure and throughput. Most of the techniques are introduced to produce reliable and efficient network without having frequent link failure during data transmission. So in future any one of the techniques on more than one routing protocol could be applied to get optimal path and best result.

## References

1. Yogendra Kumar J, Rakesh Kumar V (2012) Energy level accuracy and life time increased in mobile ad-hoc networks using OLSR. Int J Adv Res Comput Sci Softw Eng 2:7
2. Ajay K, Sheethal MS, Shany J, Priya P (2012) Optimum route life time prediction of trusted dynamic mobile nodes in large scale MANETs. Int J Ad hoc Sens Ubiquitous Comput 3
3. Kaur R, Rai MK (2012) A novel review on routing protocols in MANETs. UARJ 1(1). ISSN: 2278-1129
4. Nicklas B (1999) Zone routing protocol (ZRP). Networking Laboratory, Helsinki University of Technology, Finland
5. Johnson DB (1994) Routing in ad hoc networks of mobile hosts. In: Proceedings of the IEEE workshop on mobile computing systems and applications
6. Sun J-Z (2001) Mobile ad hoc networking: an essential technology for ervasive computing. In: International conference on info-tech and info-net, vol 3, pp 316–321
7. Sinha S, Sen S (2012) Effect of varying node density and routing zone radius in ZRP: a simulation based approach. IJCSE 4(06). ISSN: 0975-3397
8. Kaur H, Kaur S (2017) Review: Routing protocols and optimization algorithms in MANET. Int J Comput Sci Mob Comput
9. Harrag N, Refoufi A, Harrag A (2018) PSO-IZRP: new enhanced zone routing protocol based on PSO independent zone radius estimation. Wiley, Chichester
10. Kumar A, Sharma S (2018) Zone routing protocol & its enhancement technique: a review. Int J Eng Trends Technol 3
11. Priyadharshini C, Selvan D (2016) PSO based dynamic route recovery protocol for predicting route lifetime and maximizing network lifetime in MANET. In: IEEE international conference on technological innovation in ICT
12. Jagadev N, Pattanayak BK (2019) Power aware routing for MANET using PSO. IJITEE 8
13. Khan NR, Sharma S, Patheja PS (2018) Energy-aware multipath routing scheme based on particle swarm optimization (EMPSO). IRJET 5
14. Kaiwartya O, Kumar S (2014) Geocasting in vehicular adhoc networks using particle swarm optimization. ISDOC
15. Kaur H, Prabahakar G (2016) An advanced clustering scheme for wireless sensor networks using particle swarm optimization. IEEE

# Industrial IoT: Challenges and Mitigation Policies


**Pankaj Kumar, Amit Singh, and Aritro Sengupta**

**Abstract** In today's world, with the innovations of various technologies and the subsequent growth and competition in various sectors, technologies such as IoT are playing a key role in amalgamating technology and industries. This vision to use IoT in the industry is to complement the needs of industries, increase automation, receive feedback, timely response, and most importantly increase production by many folds. This drastic change and implementation of technology helped to generate more revenue but the threat of cyberattacks is still prevalent and is largely overlooked. Industrial IoT which does not follow any standardized security protocols is heavily vulnerable to cyberthreats and critical sectors like power plants, smart meters, etc., which uses IoT to connect are at high risk. In this paper, a basic overview of the cybersecurity issues in industrial IoT and various cyberattacks related to different tiers of IoT architecture is reviewed. Also, research work has discussed various mechanisms that can be used to thwart and mitigate cyberattacks on the industrial IoT systems.

**Keywords** Cybersecurity · Internet of things (IoT) · Industrial Internet of things (IIoT) · Cyberchallenges · Common vulnerabilities and exposures (CVE) · Vulnerability

## 1 Introduction

With rapid technological growth in the industrial sector, technologies such as the Internet of things, artificial intelligence (AI), blockchain technology, 5G, and cloud computing play a vital role in connecting industrial processes throughout the network.

Industrial IoT (IIoT) is rapidly becoming a reality. In critical applications, the use of intelligent network-connected devices and data management has improved productivity, optimization, and reliable communication and control. As described by Industrial Internet Consortium (IIC) [1], IIoT will significantly improve the power to

P. Kumar (✉) · A. Singh · A. Sengupta
Ministry of Electronics and Information Technology, Government of India, Delhi, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_13

take decisions, actions and connect multiple autonomous industrial control systems. Also, data from smart objects are being used to provide intelligence for decision making using big data analysis.

According to Gartner Inc.'s 2019 [2] report, 5.8 billion automotive and enterprise IoT systems will hit the market by the end of 2020, up to 21% from 2019. Building automation and connected lighting devices will have a high growth rate of around 42%, followed by the automotive industry (31%) and the healthcare sector (29%) in 2020. The emergence of networked IoT devices in the industry promotes innovation and improves efficiency and productivity. But the industrial Internet of things (IIoT) also faces cybersecurity issues/challenges that industries and IT professionals need to address.

An IoT threat report published by Palo Alto Networks [3] observed that 98% of all IoT network traffic is raw and does not have encryption, thereby revealing sensitive and confidential data of companies and consumers. Fifty-seven (57) percent of IoT devices are vulnerable to cyberattacks due to the low patch level of IoT systems. Attackers may exploit known vulnerabilities and password attacks via default device passwords. Eighty-three percent of medical imaging devices run on unsupported operating systems that may be exploited by vulnerabilities to exfiltrate patient data stored on these devices and disrupt patient care quality.

On September 20, 2016, a massive DDoS attack (around 6 Gbps traffic) was recognized on Brian Krebs' security blog (krebsonsecurity.com). It was an IoT botnet driven by Mirai malware. It scanned the Internet continuously for vulnerable IoT devices, which were then infected and used in a botnet attack. It was claimed that around 3.80 lakhs IoT devices were affected by the Mirai malware in the attack on Krebs' Web site. The affected IoT devices were mainly home routers, digital video recorders, network-enabled cameras, etc. [4]. Another advanced variant of the Mirai botnet, termed as Torii botnet, used new payloads to exfiltrate sensitive information instead of usual DDoS attacks. This botnet attack affected a wide range of devices and targeted several architectures, including MIPS, ARM, x86, x64, PowerPC, etc. [5].

In February 2019, Hoya, a Japanese Optics manufacturer, was target by a cyberattack that interrupted its production lines in Thailand. In March 2019, the Norwegian metallurgical company, Norsko Hydro, was attacked by LockerGoga Ransomware. This Ransomware attack disrupted the company's operations [6].

These attacks discussed above demonstrate that a lot of IoT devices suffer from multiple vulnerabilities. They do not have the requisite protection and protocols. Most devices have a lack of authentication measures, default login passwords, communication in open text format over the Internet, including insecure interfaces and vulnerable firmware.

The authors have reviewed several research articles [7–12] that discussed limited cybersecurity issues and mitigations techniques. However, in this paper, authors have provided a comprehensive and detailed overview of the cybersecurity risk faced by IIoT in recent years along with their related CVEs, detection, prevention, and latest mitigation techniques that ensure IoT-based industry's protection. The paper also aims to cover the latest cyberattacks trends and challenges faced by industries and end users.

The rest of the paper is structured as: Sect. 2 gives a brief description of IoT, IIoT, and their architectures. In Sect. 3, the authors have discussed different vulnerabilities and potential security threats and issues related to IIoT. In Sect. 4, the authors have discussed the best security practices for IoT application developers, industry, and end users. In Sect. 5, the authors have discussed future directions and concluded the paper.

## 2 Internet of Things (IoT) and Industrial Internet of Things (IIoT)

### *2.1 Internet of Things (IoT)*

Basically, it is a combination of static and/or mobile devices that are interconnected. These devices are being equipped with sensors, actuators, and other measuring modules which are connected to the Internet. The main purpose of IoT is to connect billions of smart objects that can sense the surrounding environment, transmit and process the data and then feedback to the environment. It uses various kinds of communication technologies and products such as cellular devices, Ethernet, Wi-Fi, Zigbee, satellite and has machine-to-machine (M2M) capabilities.

IoT devices, coupled with industrial process/industrial communication, improve the safety and sustainability of industries and also connect the physical world to its digital counterpart (known as cyberphysical system (CPS) [7]. The different applications of IoT devices are shown in Fig. 1.

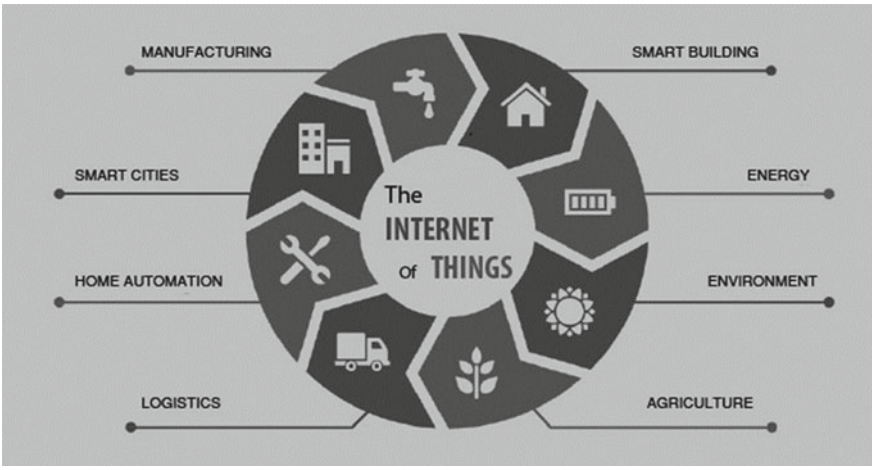



**Fig. 1** Application of IoT [13]

**Fig. 2** Benefits of IIoT


## 2.2 *Industrial IoT (IIoT)*

In the digitalized industrial world, the advent of digital and smart IoT devices aims to integrate operational technology with information technology (IT). In simple words, IIoT consists of connected smart objects, information communication technology, and cyberphysical systems. These devices provide an exchange of process, service information, real-time data by collection, processing, and communication within the industrial system, to increase overall productivity [9]. The benefits of IIoT systems have been shown in Fig. 2.

## 2.3 *IIoT Architecture*

The Industrial Internet Consortium (IIC) presented a three-tier reference architecture of IIoT that consists of edge, platform, and enterprise tier, as shown in Fig. 3.

In the edge tier, data received from different edge nodes collect at the edge gateways via a proximity network. This network connects various sensors, actuators, and control systems, which are collectively called edge nodes.

The platform tier receives the data, processes it, and transfers to the enterprise tier. It controls the processes and data flow among the tiers using an access network. It also enables management functions for devices.

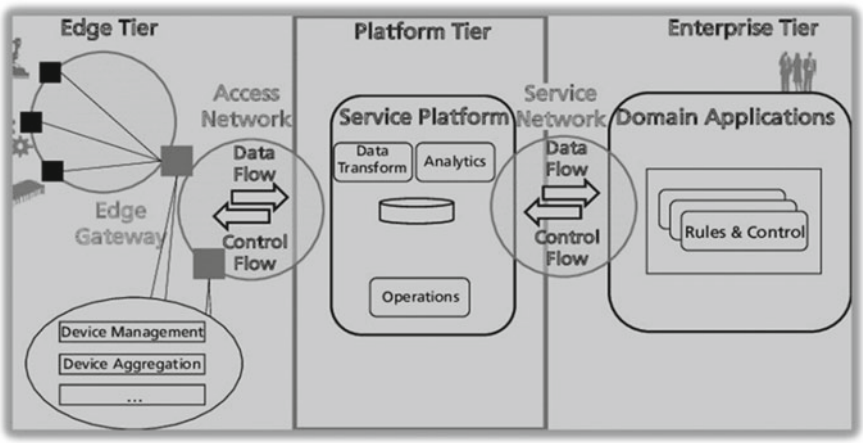



**Fig. 3** Architecture of the three-tier Industrial IoT system [10]

End users are linked via an enterprise tier to IIoT systems. This tier incorporates special frameworks and decision support systems. It also gives control command to the edge tier.

In these layers, a large amount of data is been transferred using different network topologies. In order to reduce the number of computing efforts and quantum of data, IoT devices are used at the edge nodes such as smart sensors [14].

As the number of IoT devices increases in the IIoT systems, there is a high chance of increasing cyberattacks on these devices to access valuable information. Also, the network protocols which are being used for data transfer are highly targeted to the attacks. These networks are targeted easily because their communications are not encrypted. The main attacks on IIoT systems are network-based, compromise of devices, remote code execution through malicious code, etc. To secure our IIoT ecosystem from cyberthreats, we have to establish a secure and reliable industrial environment. In the next section, the authors have discussed the potential cyberthreats and their countermeasures.

## 3 Cyberthreats in IIoT Devices

### *3.1 Cyberthreat*

A threat in cyberspace may be defined as an event in which an attacker disrupts/attacks organizational operation and its assets through denial of service, disclosure and modification of information, or unauthorized access, etc., of an industrial system.

Traditionally, cyberattacks happened at a single point of surface among hardware systems, software, or network level. But, advancement and sophistication of

cyberthreats and cyberattacker may find loopholes and bypass the security of IoT systems. To handle the dynamic nature of cyberthreat, IoT systems have to build up a defense strategy capable of handling such threats.

The IIoT systems have perimeter defense mechanisms such as firewalls and antivirus software installed within intrusion prevention/detection systems, all traffics coming from outside are intercepted and examined to ensure there are no cyberattacks from external sources into the IIoT systems. Also, a large-scale identity management system and traceback techniques may help to handle cyberthreats.

As discussed in the previous section, the cybersecurity threat exists in all the three tiers of IIoT architecture [8].

(a) Edge tier (i.e., from sensors and actuators);
(b) Platform tier (i.e., from network architecture);
(c) Enterprise tier (i.e., from user data storage).

The first tier is responsible for managing the sensors and reading data by multiple endpoints. This may allow an unauthorized attacker to add a node, fake input data, and read sensitive data of the system. This is commonly known as fake node vulnerability. The attacker can even attempt to snoop between two nodes and capture authentic information from the sender and reuse the same in the future sessions. This is known as a replay attack. The attacker can also block RFID tags or sensors that result in the loss of confidential data or manipulate the data. Also, there is a chance of cloning/tampering of RFID tags to vandalize the edge tier.

The platform tier manages the transfer of data over a wireless communication network. This wireless communication network depends on different platforms such as cloud, Bluetooth (Zigbee), LTE, and Wi-Fi. There are many concerns related to using these technologies. For instance, in Wi-Fi, there are 10–15 devices connected. If the platform tier uses one of such technologies without sufficient protection/security measures, the attacker can easily target the network and steal or manipulate the sensitive data. Also, it can suffer from DoS and MITM attacks. A DoS attack is an attack to prevent registered users from getting access to devices. This type of attack is done by flooding devices with unwanted and irrelevant requests. This jams the bandwidth and makes registered users from getting access to the devices. An attack where an attacker intercepts and alters the data sent between two nodes is known as man-in-the-middle attack [11]. This is a serious issue as far as cybersecurity is concerned, as it may allow the attacker to capture and modify data in real time.

The enterprise tier deals with the end users. There are many unlawful activities and cyberattack techniques used by the attacker to manipulate or damage the end user's sensitive information/data, for example, by injecting malicious code into the system, an attacker can shut down the complete industrial IoT infrastructure. The attacker can also make the end user access some malicious programming to gain useful information and alter it to malfunction the devices or deny its services. This tier also suffers from cross-site scripting requests. This allows an adversary to inject a script such as JavaScript. Subsequently, the adversary may modify the application and use confidential data and sensitive information illegally.

From the above discussion, it is evident that several vulnerabilities related to the architecture of IIoT systems might have cybersecurity challenges/issues.

The classification of various industrial devices/systems based on vulnerabilities and their vendors has been shown in Table 1. It has also given a number of IIoT devices that may be vulnerable to cyberattacks.

## 4 Remedial Measures and Best Practices to Mitigate Cyberthreats

Considering the risks and threats already identified, it is clear that safety measures need to be established to mitigate certain vulnerabilities emerging in IIoT area. These best practices need to be followed at all stages from development and operations to implementation and maintenance of IoT technologies.

Security measures and best practices to mitigate threats, vulnerabilities, and defined risks that affect IIoT devices and environments are categorized and summarized below.

### *4.1 Security Practices for IoT Devices Developers*

- Secure software development methodology should be followed in which a step-by-step approach to building secure software from the commencement of the development all the way to its implementation, testing, and deployment.
- Open-source software should be used with care. While selecting the open-source software for development, there is a need to consider the activity level of the community for each open-source component. An obscure and inactive open-source software might not be supported and security issues are not likely to be patched.
- Secure boot and inbuilt cryptographic protocols should be implemented. This will disallow unauthorized access and modification of the boot sequence by executing malicious code.
- Identity-based authentication, attribute-based encryption, and mutual authentication techniques will prevent attacks such as man in the middle (MITM), replay, forgery, impersonation, and eavesdropping.
- Each application should have proper authentication and access control mechanisms to provide security and integrity of the application. The application should include role-based access control for multiuser environments.
- In addition to the role, access control policy should also need to consider geolocation, device type, department, firmware version, date and time, attribute, user type to manage security issues related to unauthorized access.

**Table 1** IIoT system, its vulnerabilities, vendors, vulnerable systems, and related CVEs [5, 15–24]

| Industrial devices/software | Types of vulnerabilities, which may affect the devices/software | Layer | Vendors | No. of devices/interface at the Internet, which might be vulnerable (approx.) | Related CVEs | Solution |
|---|---|---|---|---|---|---|
| Human–machine interface programming software | 1. Memory corruption<br>2. Credential management<br>3. Lack of authentication/authorization<br>4. Cross-site request forgery<br>5. Code Injection<br>6. Heap-based buffer overflow | Platform tier | Siemens, Red lion Controls, Advantech, GE, Rockwell Automation, etc | 1400 | SSA-487246, CVE-2020-16207/11/13/15/29, CVE-2019-10978, CVE-2019-10984 | 1. Update the software with latest version<br>2. Update antivirus |
| Webcam | 1. Unauthorized access<br>2. Remote code execution<br>3. DoS overflow memory corruption<br>4. Cross-site scripting<br>5. Hard-coded credentials | Edge Tier | Cisco, Axis, Shenzhen Neo Electronics, Panasonic, Seyeon | 150,000–200,000 | CVE-2019-11219/20, CVE-2018-10664/63/62/61/59, CVE-2015-887/88, CVE-2014-8755, CVE-2006-3604 | 1. Use strong password<br>2. Use secured network<br>3. Software should be updated |

(continued)

**Table 1** (continued)

| Industrial devices/software | Types of vulnerabilities, which may affect the devices/software | Layer | Vendors | No. of devices/interface at the Internet, which might be vulnerable (approx.) | Related CVEs | Solution |
|---|---|---|---|---|---|---|
| SCADA system | 1. Inadequate encryption strength<br>2. SQL injection<br>3. Cross-site scripting<br>4. Backdoor access<br>5. Command injection and parameter manipulation<br>6. Remote code execution<br>7. Buffer error | Edge tier | Open Enterprises, Laquisscada Schneider Electric, Advantech, Delta Industrial Automation, wecon, circontrol | 600,000–700,000 | CVE-2020-6970, CVE-2019-6823/24, CVE-2018-12634/35, CVE-2019-10980/94 | 1. Update the software with latest version<br>2. Use VPN network<br>3. Communication between devices must be encrypted |

(continued)

**Table 1** (continued)

| Industrial devices/software | Types of vulnerabilities, which may affect the devices/software | Layer | Vendors | No. of devices/interface at the Internet, which might be vulnerable (approx.) | Related CVEs | Solution |
|---|---|---|---|---|---|---|
| Industrial Ethernet switches/routers | 1. Buffer overflow<br>2. MAC flooding<br>3. DHCP/ARP spoofing<br>4. SNMP vulnerability<br>5. Remote code execution | Platform tier | Siemens, Cisco, Netgear, Fortinet | 10,000–20,000 | CVE-2020-8597, CVE-2018-18065, CVE-2020-3205/3198 | 1. In routers passwords must be enabled at both the login mode and the privileged mode<br>2. Login mode passwords on Console, AUX, and VTY (telnet/ssh) interfaces must be apllied<br>3. Protect router/switches with a firewall and ACL |

(continued)

**Table 1** (continued)

| Industrial devices/software | Types of vulnerabilities, which may affect the devices/software | Layer | Vendors | No. of devices/interface at the Internet, which might be vulnerable (approx.) | Related CVEs | Solution |
|---|---|---|---|---|---|---|
| RFID Reader | 1. Time-of-check time-of-use (TOCTOU) race condition<br>2. Eavesdropping & Replay<br>3. Man-in-the-middle attack<br>4. Cloning or spoofing<br>5. Clickjacking | Edge tier | Siemens, ABUS Secvest, Impinj Speedway | 1000–2000 | CVE-2019-15126, CVE-2019-9861, CVE-2018-5304 | 1. Tag data must be locked permanently<br>2. Single-source password protection<br>3. Multipoint password protection |
| Process control system | 1. Unquoted search path<br>2. Remote code execution<br>3. Lack of encryption | Edge tier | Siemens, Cisco | 15,000–30,000 | CVE-2020-7580 | 1. Update the software with latest version<br>2. Use VPN network<br>3. Communication between devices must be encrypted. |

(continued)

**Table 1** (continued)

| Industrial devices/software | Types of vulnerabilities, which may affect the devices/software | Layer | Vendors | No. of devices/interface at the Internet, which might be vulnerable (approx.) | Related CVEs | Solution |
|---|---|---|---|---|---|---|
| Measuring and monitoring system | 1. Out-of-bounds read<br>2. Missing authentication for critical function<br>3. Lack of encryption<br>4. Use of password hash with insufficient computational effort<br>5. Cross-site scripting<br>6. Classic buffer overflow<br>7. Authentication bypass by capture–replay | Enterprise tier | Siemens, Cisco | 5000–10,000 | CVE-2020-10037/38/39/40/41/42/43/44/45 | 1. Update the software with the latest version<br>2. Use VPN network<br>3. Communication between devices must be encrypted<br>4. Update antivirus |

(continued)

**Table 1** (continued)

| Industrial devices/software | Types of vulnerabilities, which may affect the devices/software | Layer | Vendors | No. of devices/interface at the Internet, which might be vulnerable (approx.) | Related CVEs | Solution |
|---|---|---|---|---|---|---|
| Industrial real-time (IRT) device | 1. Improper input validation<br>2. Remote code execution | Edge tier | Siemens, Erricson | 10,000–20,000 | CVE-2019-10923 | 1. Update the firmware<br>2. Control system networks and remote devices locate behind firewalls and isolate them from the office network<br>3. Enable access protection and change default credentials for SNMP service |

- Device-to-device authentication may be enforced to prevent masquerading of devices by malicious parties and also to provide accountability and forensic analysis of devices.
- Secure key management includes key generation, updation, revocation, and storage of keys to prevent masquerading and device compromises attacks.
- Self-encrypting devices should be preferred to prevent unauthorized disclosure of data.

## *4.2 Best Security Practices for Communication Network*

- Restrict unauthorized access to the network and implement authentication, authorization, and accounting (AAA) systems.
- A pairing protocol may be enforced between the end users and devices to authenticate the communication. In this system, there is no need to share any prior key/password.
- A cryptographic key exchange may be used between the end users and the IoT devices through an auxiliary or out-of-band (OOB) channel to authenticate the communication.
- Implement appropriate access control lists (ACLs) for IP addresses and/or port filters.

## *4.3 Best Security Practices at the Industry Level*

- The perimeter devices should be hardened properly with the use of firewalls, intrusion prevention system (IPS), and demilitarized zone (DMZ), wherever it is necessary.
- Use tamper-proof and trusted/ secure hardware that integrates security at a different level and provide encryption and anonymity.
- To prevent the remote access attack, penetration of the firewall, and routers/switches, a secure virtual private network (VPN) should be used.
- There should be a mechanism that detects vulnerabilities, develops patches to plug the same, and applies patches to the end devices with customer consent.
- There is a need for secure device firmware that could be signed by an irreversible device identifier and securely delivered to the machine over a secure communication channel [25].
- Physical access should be restricted to critical IIoT control systems so that an unauthorized person would not have access to the ICS and safety controllers, peripheral devices, and safety networks.
- All kinds of data exchange with the isolated network such as DVDs and USB drives should be scanned before use in the terminals connected to these networks.

- Physical and logical network separation, segregation, and overlay of security zones and conduits model should be implemented.
- Devices should be updated regularly with new security patches (security at the device level), the latest firewall signatures of new malware (network level), monitoring, and analyzing log files (plant level) [8].
- Standard operating procedures (SOP) should be followed that support the IIoT policies.
- Physical and logical inventory of connected systems should be maintained.
- Integrity measurement architecture (IMA) can be used to devise mechanisms to address the issues with accidental and malicious modification of files [16].
- Hardware/software encryption schemes can be implemented to validate devices, sensors, and other components.
- Vulnerability assessment and penetration testing of IoT-related applications and hardware should be thoroughly conducted to identify potential risks.
- For mission-critical systems, organizations should establish a mechanism and develop a policy for testing of equipment to detect backdoors in firmware code/applications to prevent the supply chain attacks.
- Risk assessment and threat modeling may be done at different levels to identify all possible threats and vulnerabilities of IoT equipment. This will help in developing a mitigation plan and security framework for the organization.

### *4.4 Best Practices for End Users*

- There is a need to create security awareness among end users and employees of the organization through workshops by manufacturers, service providers, and network operators to understand the risks arising out of misconfigured IoT devices.
- End users should follow security measures such as changing default usernames and passwords, using complex passwords, enabling account lock-out policies, and deploying updates/patches provided by the manufactures and developers to the device software.

## 5 Challenges Related to Cybersecurity

There are several cybersecurity challenges in evolving and expanding Industrial IoT networks such as:

- The high cost of devices related to security features such as firewall, IPS, and DMZ.
- Lack of skills and knowledge about how to use IoT devices in a secure wireless network.
- Cybersecurity researchers have shown an inability to predict threats proactively. We have to build the capabilities to spot potential breaches ahead of time.

- In the market, there are various IoT devices and every device has its configuration and settings that may be a problem for users.
- Lack of universal manufacturing standards/protocols and immaturity of existing standards/protocols.
- There should be a National level Regulation Authority that can design a framework for building devices.
- Industrial management should be held responsible for any data leakage as per the provisions of any regulation like the General Data Protection Regulation (GDPR).
- Non-updation of software by users.

## 6 Conclusion and Future Work

In this paper, the basics of IIoT systems, their security threats and privacy issue related to the cyber are reviewed in detail. The industrial IoT system enables the industry to ease the industrial processes, but it has to face many cybersecurity challenges that have to deal with. Research work discussed various potential cyberthreats related to the IIoT system along with the countermeasures to overcome these challenges at the developer and industry level. As the next process in the future, IIoT architecture framework can be analyzed in detail, centering on cybersecurity perspectives recognized in this review to secure IIoT ecosystem.

## References

1. Industrial Internet Consortium (2016) Industrial Internet of Things volume security framework
2. Gartner says 5.8 billion enterprise and automotive IoT endpoints will be in use in 2020. EGHAM, UK, 29 Aug 2019. https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io
3. https://iotbusinessnews.com/download/white-papers/UNIT42-IoT-Threat-Report.pdf
4. https://us-cert.cisa.gov/ncas/alerts/TA16-288A
5. https://blog.avast.com/new-torii-botnet-threat-research
6. https://ics-cert.kaspersky.com/reports/2019/09/30/threat-landscape-for-industrial-automation-systems-h1-2019/
7. Sisinni E, Saifullah A, Han S, Jennehag U, Gidlund M (2018) Industrial Internet of Things: challenges, opportunities, and directions. IEEE Trans Ind Inform 10(10)
8. Lezzi M, Lazoi M, Corallo A (2018) Cybersecurity for Industry 4.0 in the current literature: a reference framework. Comput Ind 103:97–110
9. Boyes H, Hallaq B, Cunningham J (2018) Tim Watson "The industrial internet of things (IIoT): an analysis framework". Comput Ind 101:1–12
10. Al-Gumaei K, Schubay K, Friesenz A, Heymannx S, Pieper C, Pethigk F, Schriege S, Fraunhofer IOSB-INA (2018) A survey of Internet of Things and big dataIntegrated solutions for Industrie 4.0. https://doi.org/10.1109/etfa.2018.8502484
11. Burhan M, Rehman RA, Khan B, Kim B-S (2018) IoT elements, layered architectures and security issues: a comprehensive survey. Sens J
12. Kumar, Praveen R, Smys S (2018) A novel report on architecture, protocols and applications in Internet of Things (IoT). In: 2018 2nd international conference on inventive systems and control (ICISC). IEEE, pp 1156–1161

13. https://softmedialab.com/blog/how-to-develop-an-iot-app/
14. Industrial Internet Consortium (2017) The industrial internet of things, vol 1: reference architecture
15. The SANS industrial IoT security survey (2018)
16. https://www.welivesecurity.com/2020/06/15/warning-issued-hackable-security-cameras/
17. https://cve.mitre.org/
18. https://www.zdnet.com/article/175000-iot-cameras-can-be-remotely-hacked-thanks-to-flaw-says-security-researcher/
19. https://documents.trendmicro.com/assets/wp/wp-hacker-machine-interface.pdf
20. Bugeja J, Jönsson D, Jacobsson A (2018) An investigation of vulnerabilities in smart connected cameras. In: 2018 IEEE international conference on pervasive computing and communications workshops (PerCom workshops), pp 537–542
21. https://us-cert.cisa.gov/ics/advisories/icsa-19-248-01
22. https://ics-cert.kaspersky.com/advisories/klcert-advisories/2020/03/23/klcert-20-003-remote-code-execution-on-emerson-openenterprise-scada-server-version-2-83-and-all-versions-of-openenterprise-3-1-through-3-3-3/
23. https://www.trendmicro.com/vinfo/us/security/news/vulnerabilities-and-exploits/one-flaw-too-many-vulnerabilities-in-scada-systems
24. Jang-Jaccard J, Nepal S (2014) A survey of emerging threats in cybersecurity. J Comput Syst Sci 80(5):973–993
25. docs.microsoft.com

# Eclat_RPGrowth: Finding Rare Patterns Using Vertical Mining and Rare Pattern Tree


**Sunitha Vanamala, L. Padma Sree, and S. Durga Bhavani**

**Abstract** Frequent pattern mining is one of the key research areas in the Data Mining (DM) paradigm. There are many algorithms in the literature to identify the frequent itemsets whereas research on rare pattern mining is in the burgeoning stage. Rare items are the infrequent items, where few applications like medical diagnosis, telecommunications, and false alarm detection in industries demand for rare patterns and rare associations with frequent or infrequent items sets in the database. The algorithms that are used to identify frequent items can also be used to identify rare patterns. However, such algorithms suffer from RareItemProblem. Rare Pattern Mining algorithms that are based on Apriori and FP-Growth were designed but Eclat-based rare pattern mining algorithms have not been explored. This paper proposes an Eclat-RPGrowth, algorithm to find rare patterns and the support of itemset is calculated by using intersection of BitSets for corresponding $k - 1$ itemsets. Also, this research work proposes a variant of Eclat_RPGrowth as Eclat_PRPgrowth. Both the algorithms are outperformed in execution time, and with the number of rare items generated.

**Keywords** Rare items · Rare patterns · Eclat · Prefix based pattern mining · Data mining · FP-Growth

S. Vanamala (✉)
Department of CS, TSWRDCW, Warangal East, Warangal Urban, Telangana, India

L. Padma Sree
Department of ECE, VNR Vignan Jyothi Institute of Technology and Science, Hyderabad, Telangana, India

S. Durga Bhavani
School of Information Technology, JNTUH, Hyderabad, Telangana, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_14

## 1 Introduction

Data mining is the approach used for discovering hidden patterns in large databases. Association rule mining and frequent pattern mining is one of important tasks in the data mining [1]. Association rule mining is widely used in different domains like social network analysis, mobile data mining, banking, finance, and stock market data analysis.

Itemsets are categorized as frequent and Rare Itemsets. From the literature, many data mining approaches are available to discover the frequently occurring entities. However, real-world datasets of many applications contain both frequent as well as relatively rare occurring entities. Generally, these rare patterns are pruned by frequent pattern mining algorithms in exceptional cases.

The algorithms that are used to identify frequent items can also be used to identify rare patterns. However, such algorithms suffer from RareItemProblem, i.e., if support value has been set as high then, important rare patterns may be missed and if support is set to low value, many patterns are generated and it is impossible to identify the significance of all generated itemsets or patterns.

Rare association rule mining is the process of identifying associations among rare itemsets that are having low support (rare), but occurs with high confidence, being neglected by association rule mining algorithms.

Rare items are further classified into Rare Item Itemsets, which contains only rare items, these Itemsets are referred to as perfect rare itemsets in literature, imperfect rare itemsets are the itemsets that are combinations of both frequent and Rare Items and whose support is less than the user-defined MinimumFrequentSupport.

Rare patterns are especially useful in applications such as medical diagnosis, fraudulent card usage detection, to identify weak students study behavior and impact of remedial coaching, market basket analysis of rare items such as cooker and pan purchase, like this, rare pattern can be used in many applications and also in collaborative recommendation systems, this kind of patterns plays a vital role to improve sales and profit. Another example, to detect errors in a manufacturing industry like if {Fire = true} is identified as frequent, but {Fire = true, Alarm = Ringing} is rare, indicates that an alarm system is malfunctioning or fault in a device, which is being neglected by all frequent itemset mining algorithms as the stated scenario gives low support.

The objective of the proposed work is to identify perfect rare itemsets and rare item itemsets. The outline of the paper is as follows- Related work is discussed in Sect. 2. Problem formulation and description of algorithm are described in Sects. 3–5. Experimental results and discussion are given in Sect. 6. Section 7 presents the conclusion and future scope of the work.

## 2 Related Work

In literature, many researchers contributed to find frequent patterns from the database. Basically, data of Transaction Database can be represented in two formats, namely (1) *Horizontal data format*. (2) *Vertical data format*.

Algorithms to find frequent patterns that are based on horizontal data format are Apriori [1] and FP_Growth [2], these are the important and well-known algorithms, the foundation for frequent pattern mining, another algorithm Eclat [3] which is based on vertical data format is one more main algorithm of frequent pattern mining, the performance of Eclat is better when compared to Apriori, all these variations are exclusively for frequent pattern mining and can't be utilized for mining significant rare patterns. Hence, many researchers were proposed variations of Apriori, FPGrowth, and Eclat for finding rare patterns.

### *2.1 Apriori Based Rare Pattern Mining Techniques*

Apriori is based on the level-wise approach that generates $k$ length itemsets by joining $k-1$ itemsets and pruning. Algorithms with single minimum support and multiple minimum supports are the two variations proposed by various authors [4, 5] to find rare itemsets.

Apriori-Inverse proposed by Koh et al. [4] is used to mine perfectly rare itemsets, which are itemsets that only consist of itemsets below the user-specified maximum-supportthreshold (maxSup). As single minimum support is not suitable to find all rare items, authors proposed MSApriori and IMSApriori [6–8], which are two important techniques to find rare items, which are based on multiple minimum support.

Laszlo et al. [9] presented generation of rare association rules by mining of infrequent itemsets. This work describes a method for identifying rare association rules that stay hidden for regular frequent itemset mining algorithms. When compared with other methods, this method finds strong but rare associations that have local regularities in data are found. These associations are referred to as "mRI rules".

### *2.2 FP-Growth Based Rare Pattern Mining Techniques*

Apriori-based variations demonstrated to be ineffective while mining rare item patterns because there will be a significant increase in the number of candidate itemsets at each level, as the rare items are also retained during the itemset generation phase. To overcome the problems of Apriori technique, various rare itemset mining algorithms have used the tree-based technique known as FP-Growth.

CFP-growth [6] and CFP-growth++ [7], an optimized version of [6], these algorithms read the dataset once to build the CFP-tree. Then, they rebuild the CFP tree

using pruning and merging techniques. The tree reconstruction phase in these two algorithms is an expensive step in terms of computation cost, memory usage, and time consumption. Darrab et al. [8] was proposed a technique MISFP-growth, to find frequent itemsets and rare items with MIS. The tree construction in MISFP-growth is efficient.

Ashish Gupta et al. in [10] was presented pattern-growth paradigm to discover minimally infrequent itemsets. It has no subset which is also infrequent. This work uses novel algorithm of IFP min for mining minimally infrequent itemsets. Then the residual tree concept has been incorporated by using a variant of the FP-Tree structure which is known as inverse FP-tree. In order to mine the minimally infrequent itemsets, optimization of Apriori algorithm is performed. Finally, the presented tree is used for mining of frequent itemset.

Tsang et al. [5] developed RP-Tree algorithm, it is an improvement over the existing algorithms. Firstly, RP-Tree, rare pattern mining algorithm based on FP-Tree, to find rare patterns. RP-Tree focuses on rare-item itemsets which generate useful interesting rare association rules and does not waste much time in looking for non-rare-item itemsets, which are to be pruned later. RP-Tree is based on FP-Growth, which is capable of generating long patterns since the task is divided into a smaller sequence of searches for short patterns.

### *2.3 Eclat Based Rare Pattern Mining Algorithms*

Vertical data format initially proposed by Zaki et al. [3], advantage of this method is that it reduces the number of database scans to one. Support of $k$ itemsets can be calculated using intersection of corresponding $(k - 1)$ TidSets.

Sunitha et al. proposed MSApriori_VDB [15] to find frequent and rare patterns, in this algorithm, it uses multiple minimum supports that are obtained based on the frequency of occurrence. Hence, the method reduces the burden of assumption of minimum support threshold. The authors sunitha et al. [16] also developed algorithms to find rare patterns from data stream using vertical mining and bitsets.

MIS-Eclat [11] Darrab, a Vertical pattern mining algorithm for multiple support-based method uses vertical representation of data to mine both frequent and rare itemsets. Although these techniques address the rare itemset problem, they suffer from huge patterns. The problem is that it identifies both frequent as well as rare patterns.

The J. A. Jusoh et al. proposed R-Eclat [12], the vertical data set based technique, is specially used for infrequent pattern generation. The approach is based on Zaki's [3] Eclat algorithm. A DFS is used to achieve a reduced representation of the transactions of the customer database. Support count of k itemsets is calculated through determining support of intersecting tid-lists of its $k - 1$ subsets. In the R-Eclat algorithm, the traditional tidset, diffset and sort-diffset variants are improved to ensure that it is applicable generating infrequent patterns. The data characteristics influence the performance of the algorithm.

All these variations are based on basic Eclat which has the following shortcomings:

(1) In Eclat, candidate itemsets are maintained based on the equivalence class and clipped by using the prior knowledge. (2) Eclat makes the itemset to be very long based on combining the two $k$ itemeset and generate $(k + 1)$ itemset. (3) Processing time will be very high due to the intersection of two different itemsets.

This paper has proposed two new algorithms Eclat_RPGrowth, Eclat_PRPGrowth, which are based on Eclat_Growth. The proposed algorithms are capable of finding rare itemsets and rare-item itemsets: Itemset having at least one rare item. The proposed algorithms have the following benefits:

1. It uses vertical data format, therefore it requires single database scan.
2. It uses Apriori-like pruning strategy based on breadth-first search technique.
3. It generates only interesting rare patterns and avoids generating non-rare-item patterns.

## 3 Vertical Mining with Rare Pattern Tree

### *3.1 Basic Concepts and Definitions*

The transaction database is a set of transactions denoted as DB $= \{T_1, T_2, T_3, T_4, \ldots, T_n\}$, where $T_1, T_2$ are the transactions and $n$ is the number of transactions in the database. Each Transaction in the database consists of $k$ items, $k <= m$, Item set $I = \{I_1, I_2, I_3, \ldots, I_m\}$, where $m$ is the distinct items present in the database.

***Support***: An itemset X $\in$ I, then support of X is defined as the fraction of the transactions in a database that contains itemset X.

***Rare Item***: An itemset X is a rare item if Supp(X) <= MinFrequentSupportThreshold (MFT) and Supp(X) > MinRareSupportThreshold (MRT).

***Frequent Item***: An itemset X is frequent item if Supp(X) > = MinFrequentSupport-Threshold (MFT).

***Perfect Rare ItemSet***: itemset X $\in$ I is called perfect rare ItemSet if all items in the X are rare items.

***Rare item itemset***: An itemset is called rare item itemset if it has at least one rare item in it.

### *3.2 Vertical Mining with BitSets*

The proposed algorithms are implemented by using vertical data format with BitSet instead of tidset. Each bitset represents transaction ids with bits, the size of bitset is equivalent to the number of transactions in database, and each bit corresponds to

**Table 1** Sample database

| Tid | Items | Tid | Items | Tid | Items |
|---|---|---|---|---|---|
| 1 | 1, 2, 3, 4, 6 | 5 | 1, 3, 5, 6 | 9 | 9, 12, 11 |
| 2 | 1, 3 | 6 | 1, 2, 4, 7 | 10 | 9, 2, 12 |
| 3 | 1, 3, 4, 6, 7 | 7 | 9, 2, 3, 12, 10 | 11 | 9, 13 |
| 4 | 1, 2, 5, 8 | 8 | 1, 9, 12 | 12 | 9 |

**Table 2** One itemsets

| One items (itemname, bitset, support) | Frequent/rare | One items (itemname, bitset, support) | Frequent/rare |
|---|---|---|---|
| [item = 1, bs = {0, 1, 2, 3, 4, 5, 7}, support = 7] | Frequent | [item = 12, bs = {6, 7, 8, 9}, support = 4] | Rare |
| [item = 9, bs = {6, 7, 8, 9, 10, 11}, support = 6] | Frequent | [item = 4, bs = {0, 2, 5}, support = 3] | Rare |
| [item = 2, bs = {0, 3, 5, 6, 9}, support = 5] | Frequent | [item = 6, bs = {0, 2, 4}, support = 3] | Rare |
| [item = 3, bs = {0, 1, 2, 4, 6}, support = 5] | Frequent | [item = 5, bs = {3, 4}, support = 2] | Rare |
| | | [item = 7, bs = {2, 5}, support = 2] | Rare |

transaction. Example consider an item {1} in Table 1, [item = 1, bs = {0, 1, 2, 3, 4, 5, 7}, support = 7]. BitSet(bs) of an item 1 represents bit indexes of true bits, a bit is set to true if the corresponding transaction number has item 1. The sample database with vertical data format in bitsets is shown in Table 2. The vertical data format with bitset optimizes the memory usage and reduces the time to perform a intersection in the calculation of support of higher (k) order itemsets from lower order $(k - 1)$ itemsets. E.g. The intersection of item 1 and item 2 from Table 2 is given as Item 1 $\cap$ Item 2 = {{itemset = {1, 2}, bs = {0, 3, 5}, support = 3}, support is equivalent to the number of true bits.

### *3.3 Structure of a Rare Pattern Tree (RP Tree)*

In the proposed algorithm, the tree structure used is similar to the structure in Eclat-Growth [17], consists of the set of levels, depending on the max transaction length or user- specified maximum pattern length.

Each level again is a collection of tree nodes. Each node is a 6 tuple—{itemset(IS), BitSet(BS), ParentNode1(P1), ParentNode2(P2), childNodesPtrList(CPtrs List), isValid(IV). The first node in the first level is the root of a tree. ParentNode1, ParentNode2 represent the references to a parent node, which belongs to

| LevelNo (Item-SetLength) | Nodes list of corresponding level | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level-0(one Itemsets) | IS | BS | P 1 | P 2 | C Ptrs List | IV | | IS | BS | P 1 | P 2 | CPtrs List | IV | … |
| Level-1(two Itemsets) | IS | BS | P 1 | P 2 | CPtrs List | IV | | IS | BS | P 1 | P 2 | CPtrs List | IV | |
| … | … | | | | | | | | | | | | | |

**Fig. 1** Rare pattern tree structure

(childNodeLevel-1) level. childNodesPtrList is the reference to its children which are part of (nodeLevel + 1) level, isValid pointer is to reduce the number of invalid candidates generation, which is used to prune unwanted candidate generation.

The tree pattern in Eclat_Growth is constructed with pointers, whereas in this proposed method, the tree is created with arrayList for nodes of each level or itemsets of the same length. Hence, the number of references used in implementation is reduced, which is memory efficient. The structure of a Rare Pattern tree is shown in Fig. 1.

## 4 Methodology of Eclat_RPGrowth

Two new algorithms are proposed based on Eclat-Growth, First one is Eclat-RPGrowth, to find rareitem itemsets and second is Eclat_PRPGrowth to find perfect rare itemsets.

### *4.1 Schematic Block Diagram of Algorithm Eclat_RPGrowth*

**Steps in proposed algorithm—Eclat-RPGrowth (RPTree Building process)**

Figure 2 shows the main functional blocks of the algorithm Eclat_RPGrowth. It has two important phases.

1. ***Generate one itemsets List***: The transaction database has to be scanned once to convert the database to vertical representation with BitSet and to find the frequency of each unique item in the database by calculating the cardinality of the bitset and sort the items in descending order of support. Ignore or delete all the items if its support is less than the MRT (items having support less than MRT are assumed as noise and pruned), the remaining items whose support > MRT are used to generate both perfect rare itemsets and rare item itemsets.

2. ***AddItemToRarePattern tree***: For each single item in the oneitemsets list obtained in previous step, call the procedure addItemToRPTree. When the first item is added, the empty tree has to be created. Remaining items are added to RPtree one

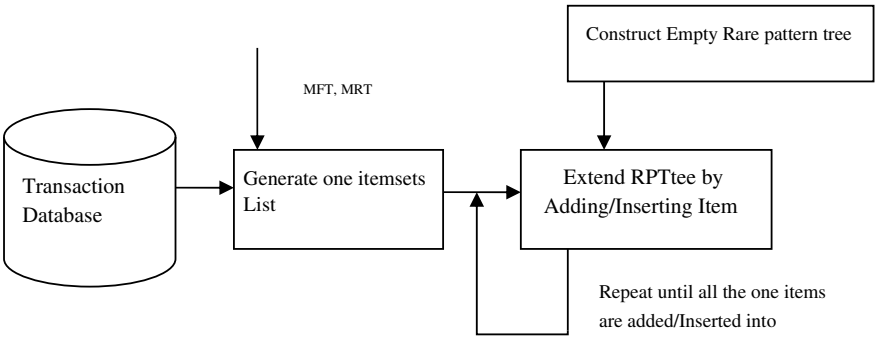



**Fig. 2** Diagram showing the major steps in the rare pattern generation

by one and the new patterns are generated row by row by combing new added node with all the nodes already present in RPTree, i.e., itemsets that are available before adding the current oneitem to the tree. Detailed step-by-step process is explained in the example section.

3. Step 2 is repeated until all the items in one itemsets list are added/inserted to tree. Traverse the tree level wise, to display rare itemsets.

**Algorithm:EclatRPGrowth**
Input: transactionfile/source file, minimumFrequentSupport(MFT), minimumRareSupport(MRT)
Output: EClat_RPTree with rare item Itemsets

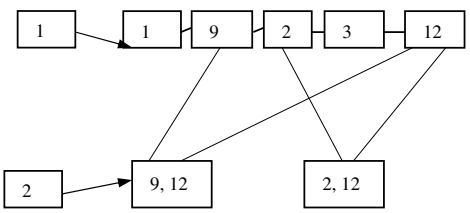
```
begin
    olist • createROneItemVMList();
     for each item sitem in olist
           begin
          if(isRare(sitem))
             begin
              addItemToEclatRPTree(sitem,MFT,MRT);
             end
           end
  end
```

```
Algorithm: addItemToEclatRPTree.
Input: Eclat-RPTree, newItem, minimumFreuentSupport, minimumRareSupport
Output: Updated Eclat-RPTree
Begin
if(eclatRPTree==null)
begin
   createEmptyTreeWithHeight(MaxTransactionLength);
end
   TreeNode newNode←newTreeNode(sItem);
   addNodetoTreeAtLevel(0,newNode);
        if(treeHeight>0)
        begin
                numLevels←treeHeight+1;
                for i in 0 to numLevels
                begin                    //i+1th level nodelist is obtained
 for each node  tempnode in currentLevel
 begin
 if(tempNode.isValid==true)
  begin
       if(isRare(tempNode)||isRare(newNode))
        begin
          newCombineNode←generateLargeItem(newNode,tempNode);
         if(isRare(newCombineNode))
            begin
            addNodetoTreeAtLevel (i+1, newCombineNode)
            else
              setAllChildrenToFalse(tempNode);
            end
          end
else
   tempNode.isValid←true;
end//else of if(tempNode.isValid==true)
end//for tempnode
end//for i
```

## 4.2 *Eclat_RPGrowth Example*

The RPTree construction procedure with the step-by-step representation is given below. Let us consider the sample transaction database in Table 1. The database has 12 transactions, each transaction has different items represented as numbers 1 to 12, the maximum transaction length in a given database is 5, hence a pattern of maximum length 5 can be detected.

**Step 1**: In the algorithm, first transactional database in Table 1 is scanned and vertical data format with bitsets is generated, all the items whose support less




| Node : Support | True BitSet indexes | IsValid |
|---|---|---|
| [1]:7 | {0, 1, 2, 3, 4, 5, 7} | True |
| [9]:6 | {6, 7, 8, 9, 10, 11} | True |
| [2]:5 | {0, 3, 5, 6, 9} | True |
| [3]:5 | {0, 1, 2, 4, 6} | True |
| [12]:4 | {6, 7, 8, 9} | True |
| [9, 12]:4 | {6, 7, 8, 9} | True |
| [2, 12]:2 | {6, 9} | True |

**Fig. 3** Building process of RP Tree, after add/insertion of items 1, 9, 2, 3, 12

than the MRT are excluded as noise and the result is shown in Table 2. Parameter values for MFT is assumed as 4, MRT as 1.

**Step 2 (Add/Insert the item to the tree)**: Items in Table 2 are added to RPtree one by one as given below.

***Extending RPTree by Adding item 1 to tree***: Initially first node (item = 1) is added to tree.

**Extending RPTree by Adding item 9, 2, and 3 to tree**: algorithm generates only rare items and rareitem itemsets, therefore a frequent item cannot be combined with another frequent item, the nodes with item = 1, item = 9 both are frequent, these two frequent items can't be combined. Similarly items 2 and 3 are also frequent, as all the items in the list are frequent, these items can't be combined to generate two itemsets. So far no two itemset is generated because all nodes added are frequent; the nodes in the tree after this step are shown in Fig. 3.

**Extending RPTree by Adding item 12**: the item 12 is a rare item, so it can be combined with a frequent or rare item and then the combined node will be added to list if its calculated support is greater than MRT. Item 12 is combined with item 1, support of {1, 12} is not satisfying MRT, so it is not included in level 2 items. The parent nodes of a node {1, 12} are items 1 and 12. Hence, is Valid entry of all the children of item 1 are set to false, so that they are not combined with item 12, this is an important step and this is the candidate pruning process applied in the proposed algorithm, which reduces number of candidate items, this will improve the performance of algorithm. However, here no children are there for item 1. Similar reasoning applies for item {3, 12} combination. Itemsets {9, 12} and {2, 12} are added to list as these items are satisfying the support. The list of itemsets after this is shown in Fig. 3.

**Extending RPTree by Adding item 4**: Item 4 is combined with Item9, its support is zero, this node is not included in level 2 list and further invalid entry of children of node 9 is set to false, similarly the combination [4, 12] is also rejected, the invalid

bit of children of node 9, node 12 are set to false. The itemsets [1, 4], [2, 4], [3, 4] are generated and added to RPtree in level 2 items list. The item [4] was not combined with nodes [9, 12] and [2, 12] as the invalid bit is already set to false and when invalid entry is checked for true/false, if it is true, the tempnode (already present in tree) is combined with new newnode otherwise without combining newnode with tempnode, the inValid entry of tempnode is changed to true. The list of nodes generated after adding item 4 are depicted in Table 3.

**Extending RPTree by Adding item**s 6, 5, 7: The list of nodes generated after adding these nodes is shown in Table 4, The total number of Rare items generated after this step are 19 {1—itemsets = 5, 2—itemsets = 11, 3—itemsets = 3}. The corresponding RPTree after final item 7 is being added is shown in Fig. 4. The invalid bit values indicate whether item 7 can be combined with it or not.

**Table 3** The list of nodes in RPTree after add item 4

| Node:Support | BitSet | IsValid | Node:Support | BitSet | IsValid |
|---|---|---|---|---|---|
| [1]:7 | {0, 1, 2, 3, 4, 5, 7} | True | [9, 12]:4 | {6, 7, 8, 9} | False |
| [9]:6 | {6, 7, 8, 9, 10, 11} | True | [2, 12]:2 | {6, 9} | False |
| [2]:5 | {0, 3, 5, 6, 9} | True | [1, 4]:3 | {0,2,5} | True |
| [3]:5 | {0, 1, 2, 4, 6} | True | [2, 4]:2 | {0, 5} | True |
| [12]:4 | {6, 7, 8, 9} | True | [3, 4]:2 | {0, 2} | True |
| [4]:3 | {0, 2, 5} | True | | | |

**Table 4** The list of nodes generated after adding an item 7 to RPtree

| Node:Support | BitSet | IsValid | Node:Support | BitSet | IsValid |
|---|---|---|---|---|---|
| [1]:7 | {0, 1, 2, 3, 4, 5, 7} | True | [2, 4]:2 | {0, 5} | False |
| [9]:6 | {6, 7, 8, 9, 10, 11} | True | [3, 4]:2 | {0, 2} | False |
| [2]:5 | {0, 3, 5, 6, 9} | True | [1, 6]:3 | {0, 2, 4} | False |
| [3]:5 | {0, 1, 2, 4, 6} | True | [3, 6]:3 | {0, 2, 4} | False |
| [12]:4 | {6, 7, 8, 9} | True | [4, 6]:2 | {0, 2} | False |
| [4]:3 | {0, 2, 5} | True | [1, 5]:2 | {3,4} | True |
| [6]:3 | {0, 2,4} | True | [1, 7]:2 | {2, 5} | True |
| [5]:2 | {3,4} | True | [4, 7]:2 | {2, 5} | True |
| [7]:2 | {2,5} | True | [1, 4, 6]:3 | {0,2} | True |
| [9, 12]:4 | {6, 7, 8, 9} | False | [3, 4, 6]:3 | {0,2} | True |
| [2, 12]:2 | {6, 9} | False | [1, 4, 7]:3 | {2, 5} | True |
| [1, 4]:3 | {0, 2, 5} | True | | | |

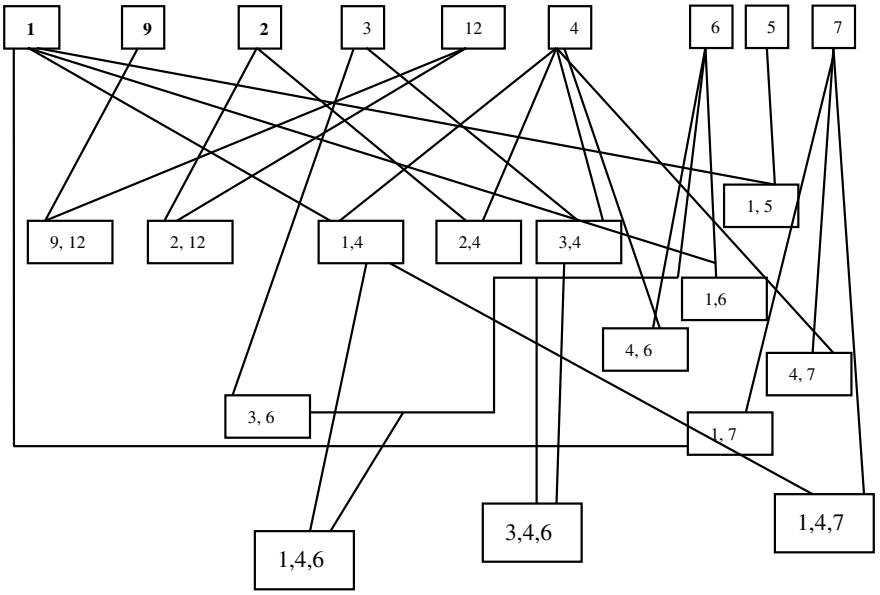



**Fig. 4** Tree structure after an item 7 is added to RPTree with 19 rare item itemsets

## 5 Methodology of EClat_PRPGrowth

The Eclat_PRPGrowth generates only perfect rare items, so the generated items are the combinations of only rare items. So when one itemsets are filtered, all the frequent one items are also filtered and retain rare items. These kinds of patterns are also useful in some critical disease diagnosis where symptoms and their co-occurrences are rare.

**Steps in proposed algorithm Eclat_PRPGrowth**

1. Scan the database to find the frequency of each unique item in the database and sort the items in descending order of support.
2. Ignore or delete all the items if its support is above the MFT or less than the MRT.
3. Create an empty tree.
4. For each single item, call the procedure addItemToRPTree.
5. Traverse the tree level wise to display rare items.

In Step 2 only rare itemsets are considered for this algorithm, the remaining process is same as EClat_RPGrowth. The list of perfect rare items generated for transactional database in Table 1 are {{12}:4, {4}:3, {6}:3,{5}:2, {7}:2, {4, 6}: 2, {4, 7}:2} with MFT as 4 and MRT as 1 and the resulting tree is shown in Fig. 5.

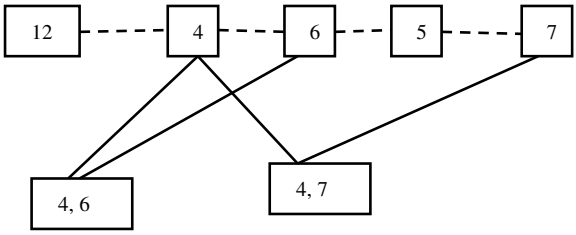



**Fig. 5** Perfect rare items generated with EClat_PRPGrowth

## 6 Experimental Results and Discussion

The proposed section shows the performance of proposed algorithms with different datasets from the UCI machine learning data mining repository [13], Datasets that are used in experiment are shown in Table 5 with their characteristics and FIMI data sets [14], tabulated in Table 6. The performance of algorithm is compared with RP-Tree and R-Eclat. The experiments were conducted on an Intel core i5 2.4 GHz machine running under the Windows 10 Operating system with 8 GB RAM.

### *6.1 Comparison of Performance*

Performance in terms of execution for Eclat-based techniques mainly depends on the number of candidate items generated and methods used to perform the intersection operation of two itemsets. In our approach, the main improvement is the good search strategy and pruning of an invalid candidate itemset generation at very early stages so

**Table 5** UCI data sets

| Dataset | Number of items | Number of transactions | Max transaction length |
|---|---|---|---|
| Teaching | 6 | 151 | 6 |
| Adult | 15 | 48,842 | 15 |
| Zoo | 18 | 100 | 18 |
| connect | 129 | 67,557 | 47 |
| Voting | 17 | 435 | 17 |
| flags | 30 | 194 | 30 |
| Dermatology | 35 | 366 | 35 |

**Table 6** FIMI data sets

| Dataset | Number of items | Number of transactions | Max transaction length | Database type |
|---|---|---|---|---|
| Mushroom | 8124 | 120 | 23 | Dense |
| Chess | 3196 | 76 | 37 | Dense |

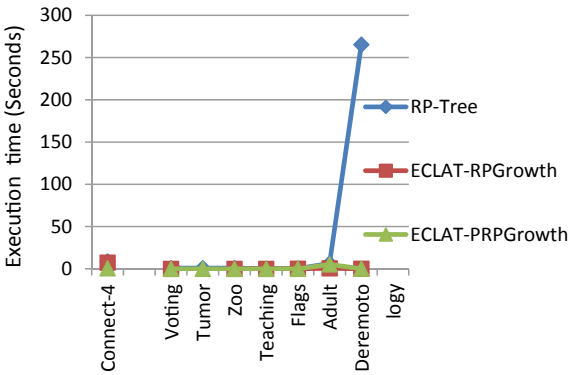



**Fig. 6** Comparison of execution time

that numbers of intersections are reduced at each level. The vertical data format uses bitset instead of tidset which reduces intersection time. The proposed algorithms have the advantages of Eclat and also the benefits of breadth-wise pruning like Apriori.

The value of MFT is assumed as 15% and absolute support for MRT as 5 to compare with RPTRee. Performance in terms execution time of three algorithms RPTree, Eclat_RPGrowth, and Eclat-PRP-Growth is shown in Fig. 5. Both the algorithms have improved in execution time compared to RPTree algorithm. A significant improvement has been detected in execution time of adult data set.

The execution time of RP-Tree and proposed algorithms is shown in Fig. 6, significant improvement can be observed in Adult dataset. The number of itemsets generated by our algorithms is also significantly reduced compared to RPTree, the comparison of number of itemsets generated is depicted in Fig. 7, and this is achieved by pruning non-rare itemsets by in the algorithms. Whereas in RPtree algorithm, generated item sets have at least one rare item, but our algorithm has been restricted at most one frequent item in generated rare itemitemsets.

Comparison of R-Eclat (Sorted diffset) with proposed algorithms is shown in Fig. 8. MFT value is assumed as 3% and MRT as 5. The proposed algorithms are better in execution time in case of both chess and mushroom data sets. Significant improvement is observed and is more than 100%. The proposed algorithms are scalable and only limitation is that it requires more memory in case of large datasets as the complete tree has to be in main memory.

## 7 Conclusion

This research work has presented a new method for finding rare itemsets in large databases using vertical mining and breadth-first rare pattern tree with two variations. This research work has utilized the two user-defined support thresholds MFT and MRT to identify rare itemsets. The first method Eclat_RPGrowth algorithm generates rare item itemsets and, the second method EclatPRP_Growth generates perfect rare

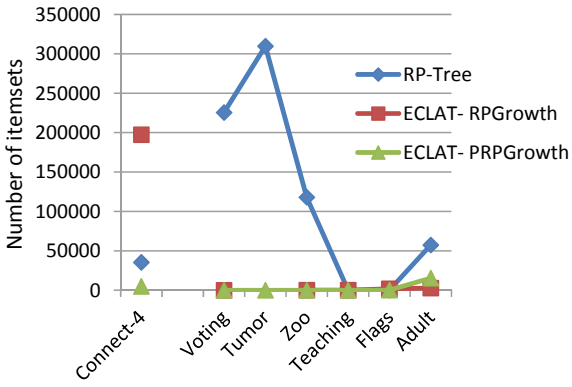



**Fig. 7** Comparison of the number of rare itemsets generated

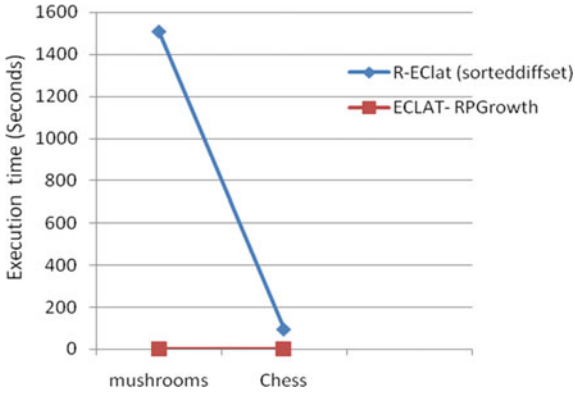



**Fig. 8** Comparison execution time of R-Eclat and Eclat-RPGrowth

items only. Two proposed methods are evaluated by comparing the performance associated with execution speed of algorithm against RPTree and R_Eclat algorithms on various datasets from the FIMI and UCI Machine learning repository. We found that in the majority of the datasets, Eclat_RPGrowth generated fewer itemsets, and execution time for our method was less compared to RPTree and R_Eclat. BitSet representation for vertical data format improved the performance of intersection operation and speed of execution. The methods are scalable and the limitation is that it demands more main memory in case of dense datasets having the large number of transactions. In the near future, it has been planned to implement these techniques for incremental databases and data streams and invent methods to generate more interesting rare patterns.

## References


1. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules in large databases. In: Proceedings of 20th international conference on very large data bases (VLDB). VLDB, pp 487–499
2. Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data, pp 1–12
3. Zaki M (2000) Scalable algorithms for association mining. IEEE Trans Knowl Data Eng 12(3):372–390
4. Koh YS, Rountree N (2005) Finding sporadic rules using apriori-inverse. In: Ho T-B, Cheung D, Liu H (eds) PAKDD 2005, vol 3518. LNCS (LNAI). Springer, Heidelberg, pp 97–106
5. Tsang S, Koh YS, Dobbie G (2011) Rp-tree: rare pattern tree mining. In: Data warehousing and knowledge discovery, Springer, pp 277–288
6. Hu YH, Chen YL (2006) Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. Decision support systems, pp 1–24
7. Kiran RU, Reddy PK (2011) Novel techniques to reduce search space in multiple minimum supports-based frequent pattern mining algorithms. In: Proceedings of the international conference on extending database technology (EDBT), pp 11–20
8. Darrab S, Ergenic B (2016) Frequent pattern mining under multiple support thresholds. In: The international conference on applied computer science (ACS), Wseas Transactions on Computer Research, pp 1–10
9. Szathmary L, Valtchev P, Napoli A (2010) Finding minimal rare itemsets and rare association rules. In: Proceedings of the 4th international conference on knowledge science, engineering and management (KSEM 2010)
10. Gupta A, Mittal A, Bhattacharya A (2011) Minimally infrequent itemset mining using pattern-growth paradigm and residual trees. In: Proceedings of the international conference on management of data (COMAD), pp 57–68
11. Darrab S, Ergenic B (2017) Vertical pattern mining algorithm for multiple support thresholds. In: International conference on knowledge based and intelligent information and engineering (KES). Procedia computer science, vol 112, pp 417–426
12. Jusoh JA et al (2018) Mining infrequent patterns using R-Eclat algorithms. J Fundam Appl Sci 24
13. Frank A, Asuncion A (2010) UCI machine learning repository. http://archive.ics.uci.edu/ml
14. Frequent itemset mining dataset repository. http://fimi.uantwerpen.be/data/
15. Vanamala S, Padma Sree L, Durga Bhavani S (2013) Efficient rare association rule mining algorithm. Int J Eng Res Appl (IJERA) 3(3):753–757
16. Ma Z, Yang J, Zhang T, Liu F (2016) An improved Eclat algorithm for mining association rules based on increased search strategy. Int J Database Theory Appl 9:251–266
17. Vanamala S, Padma Sree L, Durga Bhavani S (2014) Rare association rule mining for data stream. Int Conf Comput Commun Technol 1–6. https://doi.org/10.1109/ICCCT2.2014.7066696
18. Ma Z, Yang J, Zhang T, Liu F (2016) An improved Eclat algorithm for mining association rules based on increased search strategy. Int J Database Theory Appl 9:251–266

# Research Scholars transferring Scholarly Information through Social Medias and Networks in the Selected State Universities of Tamil Nadu


**C. Baskaran and P. Pitchaipandi**

**Abstract** The study analyzes the scholarly information transfering through Social Media and Networks by the respondents in the selected State universities of Tamil Nadu. The study examined that sharing scholarly communication on Web sites that facilitate relationship caters of needs of the researchers in any disciplines varied. This study discusses the total number of 501 respondents have reported from selected state Universities in Tamil Nadu. It determines the results of that male 260 (5.19%) and female 241 (48.1%) of the respondents from the selected eight State Universities. There are eight Universities have been chosen for collecting data from the respondents. The study also attempts to find out that Facebook users are predominant factors in terms of sharing and interacting with peer groups. The respondents highly prefer group sites (Yahoo, Google, and Whatsapp). The research analyses that social media tools for research the majority of the respondents highly preferred Facebook wall for shared the research information by the respondents in the eight Universities in Tamil Nadu.

**Keywords** Social Medias, Networks · Facebook · Twitter · Blogs · Wiki · Mendeley · Google Scholar · Research Gate

## 1 Introduction

Internet-based life is the group of online correspondences channels committed to network-based info, connection, content sharing, and coordinated effort. Sites and applications devoted to discussions, miniaturized scale blogging, social bookmarking, and wikis are among various kinds of Internet-based life. It is the best comprehended for gathering new sorts of online media, which share most or the

C. Baskaran (✉)
Librarian & Project Director (ICSSR), Alagappa University, Karaikudi 630003, India
e-mail: baskaranc@alagappauniversity.ac.in

P. Pitchaipandi
Research Associate (ICSSR), Alagappa University, Karaikudi 630003, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_15

entirety of the accompanying qualities such as Web-based life which empowers commitments and input from everybody who is intrigued [1, 2].

Most Social Media services are open to feedback and participation. They encourage voting, comments, and sharing of information. There are some barriers to access and use the content along with the password.

a. To post data about yourself as a profile
b. To post short bits of data as announcements and post photographs and documents
c. To Remark on others' substance
d. To take part in balanced and many-to-numerous discussions.

Make private or open spaces for themed dialogs long-range informal communication administrations, draw together an assortment of apparatuses, and give spaces to a scope of various gatherings to interface, thus it tends to be hard to make speculations regarding how they work. Each help is extraordinary, offers distinctive usefulness, and may be above all has its own way of life. Such societies are to a great extent the result of the individuals who are dynamic members. For instance, numerous clients of Facebook feel that the site ought to be utilized solely for social purposes. Then again, LinkedIn is utilized for the most part for proficient systems administration. In any case, societies develop and change in light of how members utilize the administration [3]. Scholastic and specialists have investigated and analyzed the numerous sides of Web-based social networking over the previous years. Associations take part in online life for the most part with the point of acquiring criticism from partners [4]. While 50 million organizations are dynamic on Facebook business pages, 2,000,000 organizations are utilizing Facebook promoting. Evidently, 88% of organizations use Twitter for promoting purposes [5].

## 2 Review of Literature

Kreps [6] reports post structuralize study by investigating how intently a person's character is reflected in their online networking profile, for example, Facebook. Garg et al. [7] examined the companion impact in an online music network and found that friends can essentially expand music disclosure. Susarla et al. [8] analyzed that video and client dataset from YouTube and found that the accomplishment of a video immensely relies upon social cooperations, which additionally decide its effect size. Gu et al. [9] explored that despite the advantages of heterophony, financial specialists are charmed by homophiles in their cooperations. Shi et al. [10] emphasized at retweet connections and find that those with feeble ties have a higher likelihood of participating in content sharing. Chiu and Huang [11] investigated the utilization of media correspondence, and the Oriesto shows that client satisfaction from informal communication locales decidedly influences their Web-based life use goal. Zaheer [12] reports the connection between the utilization of Web-based social networking among the university understudied and its consequent effect on the degree of their political interest. People have been seen as progressively dynamic in both

disconnected and online investment. Kalra and Dhingra [13] concentrated on the 46 focal university library site administrations for informal communication apparatuses classifications. When asked to the bookkeeper criticism, recommendations and remarks, contact subtleties/structure/email/telephone, library blog, informing, and so on., Sadowski et al. [14] analyzed the understudies learning advanced education organizations progressively online innovation conveyance for inside learning the executives frameworks (LMS) and outside interpersonal interaction locales means to investigate profundity exertion bunches advanced education understudies sway their instructive experience for features the requirement for steadily estimation of creative instructive activities. Mohammad and Tamimi [15] examined the exploration Imam Mohammad Ibn Saud Islamic University which is situated in Riyadh-Saudi Arabia, and the second is the University of Jordan which is situated in Amman-Jordan University understudies recognitions in regard to utilizing interpersonal interaction sites inside their learning procedure. Adewoyin et al. [16] explored the connection between Internet-based life use and administrations by administrators in government Universities in South-West Nigeria. A large number of the library benefits in university libraries in Nigeria are conveyed physically, and the conveyance of these administrations through conventional methods has been bulky and time squandering. Asnafi and Rahmani [17] surveyed the job of eExploration door in the improvement of logical insightful exercises among employees of University of Tehran's Designing School. The convenience of exploration entryway capacities, which were extricated utilizing writing survey and exploration door site, demonstrated that these abilities influenced on accomplishing researchers' examination objectives. Muscanell and Utz [18] broke down the exploration door (RG) is an Internet-based life stage for researchers, scatter their work, and fabricate their notorieties of the examination. The researchers are transferring and sharing original copies, introductions, and undertaking related materials, and asking and noting research-related inquiries. Baskaran [19] inrestricted the utilization of informal organization and media by the exploration researchers in chosen universities of Tamil Nadu. The most noteworthy 87 (31.9%) of respondents from Annamalai University and (16.8%) and (15%) of the male and female individually out of 273 complete respondents reacted to the examination. The study finds for more respondents (17.2%) of them 26–35. It is followed by (27.1%) of respondents of the executives/trade. The greater part (83.5%) of respondents utilized "Whatsapp". Baskaran [20] investigated the dominant part 73 (32.0%) of the respondents recorded from Alagappa University, though 44 (19.3%) were male and 29 (12.7%) female. The examination exposures that more respondents are 54 (18.8%) of respondents in the age bunch between 26 and 35 at Alagappa University. It is seen that out of 228, respondents were reacted from four Universities, among them significant bit 146 (64.0%) of respondents is unmarried when contrasted with (36.0%) of them wedded class. Wakefield and Wakefield [21] examined the Facebook and Twitter to show that fervor joined with enthusiasm goes about as an ideal factor for expanded online networking commitment. Xu et al. [22] explored the picture and good convictions joined with network arrangements and friend pressure goes about as hindrances to hostility via Web-based networking media. Baskaran [23] investigated those 11,941 records on informal organizations and media recovered from the

Web of Science database during the time of the study. The investigation found that more than ten distributions were contributed by an individual region out of 11,941 records during the period. Liu Y has contributed 37 (0.31%) of the distributions as a top positioned originator in the exploration.

## 3 Objectives of the Study

1. To know the Gender-wise respondents of concentrate on use social medias and networks in the state universities of Tamil Nadu.
2. To discover the utilization of SNS/Media on the modules for sharing research information by the respondents.
3. To examine the respondents explored towards preference of different type of social Medias in the selected State Universities.
4. To evaluate the exploration on the preference of social Medias tools by the respondents.
5. To analyze the respondent preference of social Medias Research Citration Indexes (RCI) in the selected State Universities of Tamilnadu.

## 4 Methodology

The examination endeavors to measure the information by appropriation and gathered from the respondents in the selected state Universities in Tamil Nadu. The present investigation incorporates only full-time Ph.D. and examines researchers engaged from chosen state universities in Tamil Nadu. There were eight state universities for conveying the survey and gathering the information. The eight Universities have been accredited with A and A+ by NAAC. The eight universities are University of Madras, Annamalai University, Bharathiar University, Bharathidasan University, Madurai Kamaraj University, Alagappa University, Manonmaniam Sundaranar University, and Periyar University. An aggregate of 520 total questionnaire disttributed among respondents in the eight universities in Tamil Nadu. Further, total number of 501 (96.34%) of the respondents is taken the survey from Social Science Departments in the selected Universities. The investigation takes place for the reason to focuss the carries of the factual dissects with ‘SPSS 20.0’ for Windows to explored and condense information gathered from the respondents. Further, the endeavor to investigate the rate instrument is utilized to discover most of the utility on specific factors, chi-square test, and one way ‘ANOVA’ tests were utilized for analyzing the data.

## 5 Analysis and Discussion

Table 1 indicated that the respondents were from state universities that are accredited "A" Grade by "NAAC". A total number of 501 respondents were reported in this study. The study analyzed that majority 87 (17.4%) of the respondent are from Annamalai University, whereas 46 (9.2%) of the male and 41 (8.2%) of the female. It can be noticed that more than 14% of the respondents are from Alagappa University (14.6%) and Periyar University (14%), and it followed by Manonmaniam Sundaranar University (13.2%), Bharathiar University (12.4%), Bharathidasan University (10.2%), and Madurai Kamaraj University (10%). The study could notice that very least respondent reported from the University of Madras (8.4%) out of eight Universities in Tamil Nadu. It is concluded that out of 501 respondents 260 (51.9%) of the male, 17.8% of less than that male respondents, the female participants witnessed 241 (48.1%) of them responded from eight state universities in Tamil Nadu (Fig. 1).

The study found from Table 2 that more than 14% of the respondents are from Alagappa University (14.6%) and Periyar University (14%). It was followed by Manonmaniam Sundaranar University (13.2%), Bharathiar University (12.4%), Bharathidasan University (10.2%), and Madurai Kamaraj University (10%). The study could notice that very least respondent reported from the University of Madras (8.4%) out of eight universities in Tamil Nadu. It is concluded that out of 501 respondents 260 (51.9%) of the male, 17.8% of less than that male respondents, and the female participants witnessed 241 (48.1%) of them responded from eight state universities in Tamil Nadu (Fig. 2).

1. **Facebook**: The respondents reported that they used SNs in the selected universities in Tamil Nadu. The study can be witnessed that majority 52 (10.4%) of them used Facebook from Annamalai University and Bharathidasan University. It followed by 50 (10%) of the respondents used Social Networks and Medias from Periyar University.

**Table 1** Gender-wise respondents of the state universities of Tamil Nadu

| S. No. | Name of the University | No. of the respondents | | Total |
|---|---|---|---|---|
| | | Male | Female | |
| 1 | Alagappa University | 44 (8.8) | 29 (5.8) | 73 (14.6) |
| 2 | Bharathidasan University | 29 (5.8) | 22 (4.4) | 51 (10.2) |
| 3 | Bharathiar University | 27 (5.4) | 35 (7.0) | 62 (12.4) |
| 4 | University of Madras | 23 (4.6) | 19 (3.8) | 42 (8.4) |
| 5 | Periyar University | 29 (5.8) | 41 (8.2) | 70 (14.0) |
| 6 | Annamalai University | 46 (9.2) | 41 (8.2) | 87 (17.4) |
| 7 | Madurai Kamaraj University | 33 (6.6) | 17 (3.4) | 50 (10.0) |
| 8 | Manonmaniam Sundaranar University | 29 (5.8) | 37 (7.4) | 66 (13.2) |
| | Total | 260 (51.9) | 241 (48.1) | 501 (100.0) |

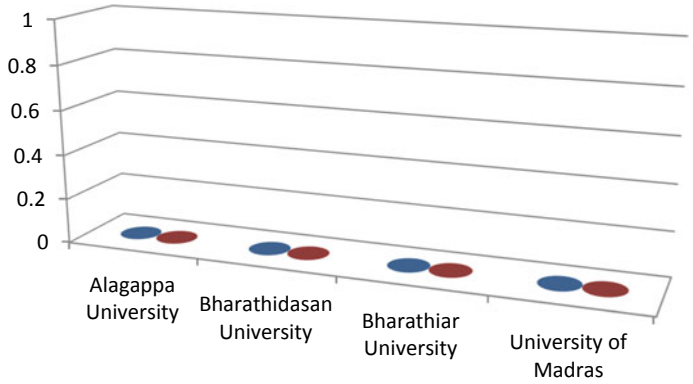



**Fig. 1** Gender-wise respondents of the state universities of Tamil Nadu

2. **Twitter**: The majority 21 (4.2%) of the respondents from Alagappa University and Bharathiar University used "Twitter". It has been seen that rest of 103 (20.1) of them utilized "Twitter" by the respondents from Madurai Kamaraj University, Annamalai University, Bharathidasan University, University of Madras, and Manonmaniam Sundaranar University.
3. **YouTube**: The majority 56 (11.2%) of them reported them used from Periyar University, 10.6% of them used from Alagappa University, and 10.2 and 9.8% of them used by the respondents from Bharathiar University and Annamalai University. It is witnessed that
4. **Tumblr/Messenger**: The majority of them reported that "Not use" the social networks of which 45.4% of them higher than predominantly use Tumblr/Messenger. Therefore, this Tumblr/Messenger is a not popular one and shares the information rarely through the social networks.
5. **Whatsapp**: The majority 86.2% of the respondents replied that they were "Use" Whatsapp since this social media is most popular among all category of people, likewise, the respondents disseminate the scholarly information sharing the scholarly contents via this. It is concluded that when to compare the "Non-use" category respondents 72.4% less than that use of Whatsapp by the respondents.
6. **Google+**: The use of Google+ social networks "Use" by 288 (57.5%) respondents is in the state universities of Tamil Nadu. It shows only 15.2% of the respondents are higher than "Non-use" respondents.
7. **Instagram**: The majority 367 (73.3%) of the respondents were not "Use" Instagram, whereas 36.6%. of the "Use" respondents are less than the "Non-use" category.
8. **Others**: Above listed seven social networks are popularly known and used by the respondents in the state universities of Tamil Nadu. The "Others" types of social media are categorized as excluded above described seven social media. It concluded that that majority 86.0% of the respondents are replied the "Non-use" category of Social Media.

**Table 2** Usage of different types of SNs among the respondents

| S. No. | Name of social medias and networks | Alagappa University | Bharathidasan University | Bharathiar University | University of Madras | Periyar University | Annamalai University | Madurai Kamaraj University | Manonmaniam Sundaranar University | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Facebook | 46 (9.2) | 40 (8.0) | 52 (10.4) | 37 (7.4) | 50 (10.0) | 52 (10.4) | 42 (8.4) | 36 (7.2) | 355 (70.9) |
| 2 | Twitter | 21 (4.2) | 16 (3.2) | 21 (4.2) | 16 (3.2) | 20 (4.0) | 18 (3.6) | 19 (3.8) | 14 (2.8) | 145 (28.9) |
| 3 | YouTube | 53 (10.6) | 37 (7.4) | 51 (10.2) | 35 (7.0) | 56 (11.2) | 49 (9.8) | 44 (8.8) | 45 (9.0) | 370 (73.9) |
| 4 | Tumblr/Messenger | 23 (4.6) | 11 (2.2) | 25 (5.0) | 13 (2.6) | 24 (4.8) | 18 (3.6) | 12 (2.4) | 11 (2.2) | 137 (27.3) |
| 5 | Whatsapp | 63 (12.6) | 45 (9.0) | 55 (11.0) | 41 (8.2) | 60 (12.0) | 67 (13.4) | 45 (9.0) | 56 (11.2) | 432 (86.2) |
| 6 | Google + | 47 (9.4) | 24 (4.8) | 41 (8.2) | 25 (5.0) | 38 (7.6) | 53 (10.6) | 28 (5.6) | 32 (6.4) | 288 (57.5) |
| 7 | Instagram | 17 (3.4) | 13 (2.6) | 27 (5.4) | 19 (3.8) | 17 (3.4) | 17 (3.4) | 15 (3.0) | 9 (1.8) | 134 (26.7) |
| 8 | Others | 18 (3.6) | 4 (0.8) | 9 (1.8) | 8 (1.6) | 8 (1.6) | 5 (1.0) | 10 (2.0) | 8 (1.8) | 70 (14.0) |

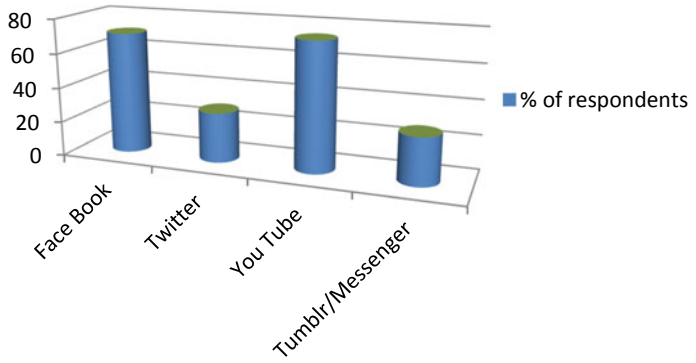



**Fig. 2** Usage of different types of SNs among the respondents

The study discussed that SNs modules used by respondents from the selected state university of Tamil Nadu. Table 3 analyzed that above 50% of the respondents, among them, 339 (67.7%) of the respondents preferred "Group sites" (e.g., Yahoo Group, Google Group and Whatsapp group, etc.), followed by 285 (56.9%) of them preferred "Instant messaging sites" (Whatsapp). The study witnessed that below (50%) of respondents preferred, out of the 149 (29.7%), 87 (17.4%) 76 (15.2%), and 63 (12.6%) of them noticed "Profile-based multi-dimensional site", "discussion forms", short messaging services, and blogs, respectively. Further, the study could be witnessed from Fig. 3, the other SNs modules preferred by the respondents to share the scholarly information, only 41 (8.2%) of them recorded for preferred besides six modules by the respondents of the study (Fig. 4).

The respondents do prefer tools for sharing contents/texts related to scholarly information share through SNs in the state universities in Tamil Nadu. Table 4 analyzed that there are seven tools as given. The majority 211 (42.1%) of the respondents prefer "Facebook walls" for sharing research contents/texts. It followed that 201 (40.1%) of them preferred "Yahoo/Google+" which is less than only 2% to "Facebook walls".

It is discussed that 157 (31.3%), 42 (22.4%), and 92 (18.4%) of them witnessed that preferred tools for sharing scholarly information on "YouTube comments", "Blogs/wiki articles," and "other tools." Further, the study can be reported that less than 10% of the respondents given their opinion for them sharing scholarly information through SNs, among them "Tweets" (128%), "Delicious Book marks" (7.6%), and "Flickr Comments and Tags" (3.3%).

The social science research scholars concentrate to publish the papers in various journals, whereas they do expect and try to get more citation counts and their h-Index. Table 5 depicts that respondents preferring social media are to share research citation indexes (RCI) in the state universities of Tamil Nadu. The study has been witnessed that there have been listed six "Research citation Index". More than 80% of the respondents among 492 (98.2%) and 431 (86.0%) stated "Get CITED" and

**Table 3** Use of SNS/medias modules by the respondents

| S. No. | Types of modules | Name of the Universities | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Alagappa University | Bharathidasan University | Bharathiar University | University of Madras | Periyar University | Annamalai University | Madurai Kamaraj University | Manonmaniam Sundaranar University | |
| 1 | Group sites (e.g., Yahoo, Google, and Whatsapp, etc.) | 47 (9.4) | 31 (6.2) | 47 (9.4) | 30 (6.0) | 52 (10.4) | 57 (11.4) | 32 (6.4) | 43 (8.6) | 339 (67.7) |
| 2 | Blogs | 8 (1.6) | 6 (1.2) | 7 (1.4) | 11 (2.2) | 10 (2.0) | 6 (1.2) | 6 (1.2) | 9 (1.8) | 63 (12.6) |
| 3 | Discussion forms | 13 (2.6) | 13 (2.6) | 13 (2.6) | 11 (2.2) | 16 (3.2) | 7 (1.4) | 6 (1.2) | 8 (1.6) | 87 (17.4) |
| 4 | Instant message sites | 43 (8.6) | 29 (5.8) | 39 (7.8) | 31 (6.2) | 41 (8.2) | 43 (8.6) | 29 (5.8) | 30 (6.0) | 285 (56.9) |
| 5 | Short message services (Twitter) | 14 (2.8) | 6 (1.2) | 11 (2.2) | 10 (2.0) | 9 (1.8) | 6 (1.2) | 10 (2.0) | 10 (2.0) | 76 (15.2) |
| 6 | Profile -based multi-dimensional sites (e.g., Friender and Facebook) | 16 (3.2) | 16 (3.2) | 27 (5.4) | 19 (3.8) | 20 (4.0) | 23 (4.6) | 13 (2.6) | 15 (3.0) | 149 (29.7) |
| 7 | Other modules | 11 (2.2) | 2 (0.4) | 3 (0.6) | 4 (0.8) | 4 (0.8) | 6 (1.2) | 5 (1.0) | 6 (1.2) | 41 (8.2) |

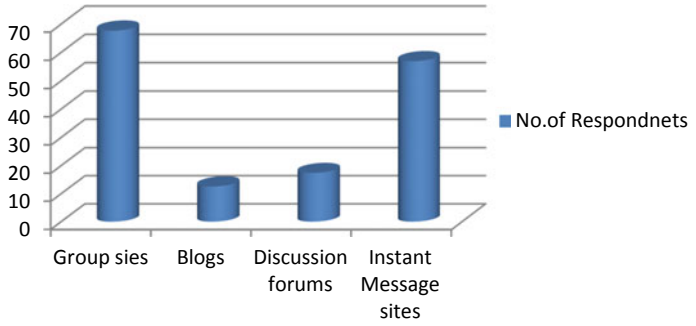



**Fig. 3** Use of SNS/Medias modules by the respondents

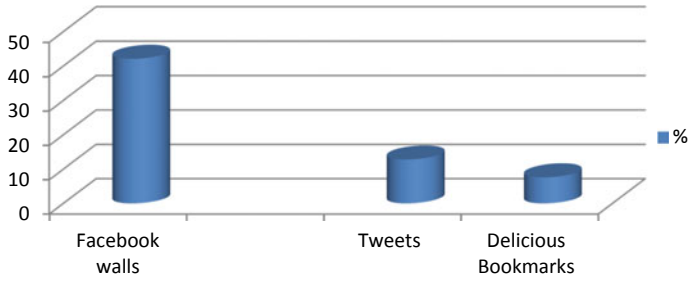



**Fig. 4** Tools preference for sharing contents and texts through SNs

"Google Scholar", respectively. More than 20% of the 146 (29.1%) and 115 (23.0%) of the respondents recorded "Scopus" and "Web of science", respectively.

Table 6 analyzed the respondents of the study in the selected state universities in Tamil Nadu. The study finds that the respondents used social media to find out research citations for the publications. The study has been witnessed from the analysis that FR is 3.973 and FP is 0.0003. It has been analyzed that "Null Hypothesis" is rejected from the above analysis. Further, the study has been discussed that there is a critical contrast between the respondents from selected state universities, and they utilized citations of research publications and subject references.

Table 7 discussed the analysis made from the study of the respondents from selected state universities in Tamil Nadu. The respondents preferred social media for other than research activities. The study has been witnessed that FR is 1.6154 and FP is 0.1286. Further, it is witnessed that "Null Hypothesis" is accepted in the analysis. It has inferred that there is no difference between the respondents from selected state universities in Tamil Nadu, and they have been used social media for other than research activities.

**Table 4** Tools preference for sharing contents and texts through SNs

| S. No. | Types of tools | Name of the Universities | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Alagappa University | Bharathidasan University | Bharathiar University | University of Madras | Periyar University | Annamalai University | Madurai Kamaraj University | Manonmaniam Sundaranar University | |
| 1 | Facebook walls | 31 (6.2) | 25 (5.0) | 38 (7.6) | 19 (3.8) | 32 (6.4) | 23 (4.6) | 29 (5.8) | 14 (2.8) | 211 (42.1) |
| 2 | Tweets | 12 (2.4) | 4 (0.8) | 10 (2.0) | 5 (1.0) | 8 (1.6) | 8 (1.6) | 9 (1.8) | 8 (1.6) | 64 (12.8) |
| 3 | Delicious Bookmarks | 5 (1.0) | 5 (1.0) | 7 (1.4) | 3 (0.6) | 4 (0.8) | 3 (0.6) | 4 (0.8) | 7 (1.4) | 38 (7.6) |
| 4 | YouTube comments | 25 (5.0) | 14 (2.8) | 23 (4.6) | 10 (2.0) | 22 (4.4) | 22 (4.4) | 28 (5.6) | 13 (2.6) | 157 (31.3) |
| 5 | Flickr comments/tags | 3 (0.6) | 1 (0.2) | 3 (0.6) | – | 2 (0.4) | 3 (0.6) | 2 (0.4) | 3 (0.6) | 17 (3.39) |
| 6 | Yahoo/Google+ | 42 (8.4) | 11 (2.2) | 20 (4.0) | 14 (2.8) | 28 (5.6) | 42 (8.4) | 17 (3.4) | 27 (5.4) | 201 (40.1) |
| 7 | Blogs/wiki articles | 14 (2.8) | 13 (2.6) | 11 (2.2) | 13 (2.6) | 14 (2.8) | 17 (3.4) | 14 (2.8) | 16 (3.2) | 112 (22.4) |
| 8 | Any others | 11 (2.2) | 7 (1.4) | 18 (3.6) | 10 (2.0) | 16 (3.2) | 8 (1.6) | 9 (1.8) | 13 (2.6) | 92 (18.4) |

**Table 5** Preference of social medias on research citation indexes (RCI)

| S. No. | Research citation indexes | Name of the universities | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Alagappa University | Bharathidasan University | Bharathiar University | University of Madras | Periyar University | Annamalai University | Madurai Kamaraj University | Manonmaniam Sundaranar University | |
| 1 | Google Scholar | 68 (13.6) | 45 (9.0) | 57 (11.4) | 36 (7.2) | 66 (13.2) | 65 (13.0) | 42 (8.4) | 52 (10.4) | 431 (86.0) |
| 2 | Cite Seer | 6 (1.2) | 1 (0.2) | 2 (0.4) | 2 (0.4) | 3 (0.6) | 1 (0.2) | 1 (0.2) | 5 (1.0) | 21 (4.2) |
| 3 | Get CITED | 1 (0.2) | 1 (0.2) | 2 (0.4) | – | 1 (0.2) | 87 (17.4) | 50 (10.0) | 62 (12.4) | 492 (98.2) |
| 4 | Math Scinet | 1 (0.2) | – | 2 (0.4) | – | 1 (0.2) | 2 (0.4) | 2 (0.4) | 1 (0.2) | 9 (1.8) |
| 5 | Scopus | 19 (3.8) | 18 (3.6) | 18 (3.6) | 15 (3.0) | 30 (6.0) | 19 (3.8) | 10 (2.0) | 17 (3.4) | 146 (29.1) |
| 6 | Web of science | 16 (3.2) | 16 (3.2) | 11 (2.2) | 10 (2.0) | 16 (3.2) | 19 (3.8) | 10 (2.0) | 17 (3.4) | 115 (23.0) |
| 7 | EBSCO | 1 (0.2) | 4 (0.8) | 7 (1.4) | 1 (0.2) | 3 (0.6) | – | 5 (1.0) | 3 (0.6) | 24 (4.8) |
| 8 | Pro Quest | 4 (0.8) | 5 (1.0) | 14 (2.8) | 6 (1.2) | 8 (1.6) | 8 (1.6) | 4 (0.8) | 4 (0.8) | 53 (10.6) |
| 9 | Others | 10 (2.0) | 3 (0.6) | 17 (3.4) | 9 (1.8) | 11 (2.2) | 18 (3.6) | 9 (1.8) | 13 (2.6) | 90 (18.0) |

**Table 6** Respondents Vs. used SNs for tweeting for research ideas

| Variable groups | D.F | Sum of squares | Mean squares | *F* ratio | *F* prob. |
|---|---|---|---|---|---|
| Between | 7 | 62.0412 | 8.8630 | 4.9909 | 0.0000 |
| Within | 493 | 875.4958 | 1.7759 | | |
| Total | 500 | 937.5369 | | | |

**Table 7** Respondents Vs. prefer SNs/medias for other than research activities

| Variable groups | D.F | Sum of squares | Mean squares | *F* ratio | *F* prob. |
|---|---|---|---|---|---|
| Between | 7 | 19.2503 | 2.7500 | 1.6154 | 0.1286 |
| Within | 493 | 839.2926 | 1.7024 | | |
| Total | 500 | 858.5429 | | | |

## 6 Conclusion

The study considered the quantitative data from the respondents in the selected state universities in Tamil Nadu. In the proposed work, the majority of respondents reported from Annamalai University. Now, considering Whatsapp and Facebook is highly preferable to share the research information with others. Likewise, the study witnessed the majority of the respondents preferred Facebook to share the research information. Presently, the researchers are more interested to count their citations on their research work, and h-Index on the purpose of Google Scholar is the primary one it can be registered easily by every researcher. The study analyzed that more respondents highly preferred Google scholar for analyses of their h-Index, citations, and research impact of the publications of social science scholars. The present context of research to an individual depends on the research information and communication disseminated by the social media which is twitted, and Mendeley on the research publications score can be counted to Altmetrics score which will be distinct to the researchers. Social media plays the most vibrant information communication sharing devices in the present context. Research on such networks ought to extend to consider the interaction among the researchers for searching needful information against reviews collecting, fact finding, and an innovative concept on the research. The researcher attempts to seek the methods of research, in terms of origin to develop and support clients "Natural Inspirations". From a hierarchical point of view, examination via Web-based networking media should move past the ordinary dyadic perspective on the connection between an online network and a firm, and the scientists get the insightful correspondences in expanding research spaces, which centers around re-conceptualizing on the Web clients as an eco-arrangement of partners. Online networking has restored the elements between associations, representatives, customers, and academic networks of the information society.

**Acknowledgements** The research project is supported by the Indian Council of Social Science Research (ICSSR), New Delhi.

## References


1. Dong JQ, Wu W (2015) Business value of social media technologies: evidence from online user innovation communities. J Strateg Inf Syst 24(2):113–127
2. Muscanell N, Utz S (2017) Social networking for scientists: an analysis on how and why academics use research gate. Online Inf Rev 41(5):744–759
3. Cann A (2011) Social media: a guide for researchers. Res Inf Netw. www.rin.ac.uk/social media-guide
4. Phang CW, Kankanhalli A, Tan BC (2015) What motivates contributors versus Lurkers? An investigation of online feedback forums. Inf Syst Res 26(4):773–779
5. Lister M (2017) 40 essential social media marketing statistics for 2017. http://www.wordstream.com/blog/ws/2017/01/05/social-media-marketing-statistics
6. Kapoor KK et al (2018) Advances in social media research: past, present and future. Inf Syst Front 20:531–558. https://doi.org/10.1007/s10796-017-9810-y
7. Zaheer L (2016) Use of social media and political participation among University students. Pakistan Vis 17(1):278–299
8. Kalra J, Dhingra S (2016) Use of social networking tools by the libraries of central universities of india: a study, scientific society of advanced research and social change SSARSC. Int J Lib Inf Netw Knowl 1(1)
9. Sadowski C, Pediaditis M, Townsend R (2017) University students' perceptions of Social Networking Sites (SNs) in their educational experiences at a regional Australian University. Australasian J Edu Technol 33(5)
10. Mohammad H, Tamimi H (2017) Students' perception of using social networking websites for educational purposes: comparison between two Arab Universities. Int J Manag Inf Technol (IJMIT) 9(2)
11. Adewoyin OO, Onuoha UD, Ikonne CN (2017) Social media use and service delivery by librarians in federal universities in the South-West. Nigeria Lib Philos Practice (e-journal). 1641
12. Asnafi AR, RahmaniMMA (2017) Knowledge and information science, utilizing research gate social network by Iranian engineering. Lib Philoso Practice (e-journal) 1585. http://digitalcommons.unl.edu/libphilprac/1585
13. Baskaran C (2018) Disseminating scholarly information access through Social Networks (SNs) and media by the social science research scholars in selected State Universities of Tamilnadu. J Adv Lib Inf Sci 8(3):124–131
14. Baskaran C (2019) Scholarly Information Share through Social Networks (SNs) and Medias among Social Science Scholars in selected State Universities in Tamilnadu. Int J Lib Inf Stud 9(3):83–92
15. Wakefield R, Wakefield K (2016) Social media network behavior: a study of user passion and affect. J Strateg Inf Syst 25(2):140–156
16. Chiu CM, Huang HY (2015) Examining the antecedents of user gratification and its effects on individuals 'social network services usage: the moderating role of habit. Eur J Inf Syst 24(4):411–430
17. Gu B, Konana P, Raghunathan R, Chen HM (2014) Research Note—the allure of homophily in social media: evidence from investor responses on virtual communities. Inf Syst Res 25(3):604–617
18. Shi Z, Rui H, Whinston AB (2014) Content sharing in a social broadcasting environment: evidence from twitter. MIS Q 38(1):123–142
19. Kreps D (2010) My social networking profile: copy, resemblance, or simulacrum? A post structuralist interpretation of social information systems. Eur J Inf Syst 19(1):104–115
20. Xu B, Xu Z, Li D (2016) Internet aggression in online communities: a contemporary deterrence perspective. Inf Syst J 26(6):641–667
21. Garg R, Smith MD, Telang R (2011) Measuring information diffusion in an online community. J Manag Inf Syst 28(2):11–38

22. Susarla A, Oh JH, Tan Y (2012) Social networks and the diffusion of user-generated content: evidence from YouTube. Inf Syst Res 23(1):23–41
23. Baskaran C (2020) Research patterns on the social networks and media: a scientometric portrait, handbook of research on emerging trends and technologies in library and information. Science. https://doi.org/10.4018/978-1-5225-9825-1.ch014

# Twitter-Based Disaster Management System Using Data Mining


**V. G. Dhanya, Minu Susan Jacob, and R. Dhanalakshmi**

**Abstract** Social media is an essential part of life for most people around. No wonder even during emergencies like flood or cyclone, more and more people look up to Twitter, Facebook, WhatsApp groups, etc., for immediate assistance. This helps to get data from even remote places and from small groups which will be difficult to reach. This sheer amount of data generated during a short span of time is also the challenge in this approach. Even when there are resources available for help, many requests could go unnoticed. This paper addresses above-mentioned problem by collecting the generated requests for help and resource availability and plot the location in the map. Request data shall be analysed using three machine learning algorithms called linear ridge regression, SGD classifier and Naive Bayes algorithm for the initial filtering and will be passed through natural language processing to match needs and offers within a given geographic boundary. The system is working with 96% accuracy for linear ridge regression and Naive Bayes classifier and 95% accuracy for SGD classifier. The report shall be published to provide a centralized status of requests. This brings more efficient management of disaster situations.

**Keywords** Twitter · Machine learning · Disaster · Datamining

## 1 Introduction

Humans are exploiting nature in many ways. Now humans are paying for that. Currently, the world is facing so many natural disasters. In 2018–19, Kerala witnessed drastic flood, and more than 100,000 people had been evacuated from their home. In

V. G. Dhanya (✉) · M. S. Jacob · R. Dhanalakshmi
KCG College of Technology, Anna University, Chennai, India
e-mail: deepa.ece@kcgcollege.com

M. S. Jacob
e-mail: minususanjacob@gmail.com

R. Dhanalakshmi
e-mail: dhanalakshmisai@gmail.com



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_16

2015, heavy flood due to heavy rain fall happened in Chennai, more than 500 people were killed, and over 18 lakh people were displaced. This all shows how natural disasters make human life difficult.

Nowadays, people depend on social media in a wide range for information sharing around the world. The social media is being considered as a medium for emergency communication because of its growing ubiquity, communications rapidity and cross-platform accessibility [1]. Twitter and Facebook are the major ones among this. Twitter is one microblogging service that allows its subscribers to broadcast short messages, called tweets, of up to 140 characters. These tweets are used to share relevant information and report news [1]. Twitter is taken here for data analysis. For recent years, several data analysis is done on Twitter than the other social media. In other social media like Facebook, messages are coming in many ways in different word length and pattern. The uniqueness of Twitter is tweets coming in structured and precise manner. So, the analysis of data is comparatively easy. Twitter provides excellent APIs for analysing and understanding data. This makes Twitter as a wide range platform for the data analyst. So many requests are coming in Twitter at the time of natural disasters in the form of help request and willing to give help for the victims. But this request is unnoticed because of the immense number of tweets coming at the time of natural disasters. This paper addresses that problem efficiently with the three machine learning algorithms. Linear ridge regression, SGD classifier and Naïve Bayes algorithm are used here, and the performance is calculated with respect to accuracy, precision and recall.

The aim of the proposed system is to identify the help request tweet and tweets coming with the information of the resource availability. So many junk tweets are coming under the same hash tag, so a data set is trained to find the required tweets. Identifying the location of the tweets is an important factor in the disaster management. Some tweets are help request for some other persons in a different location. The location with the tweets will not help here to identify the exact location. This model aims to identify such location with the help of machine learning algorithms and plot the location in the map. So, the rescue operators can easily find the location. This will improve the efficiency of the disaster management in terms of time and resources.

## 2 Related Works

Nguyen et al. [2] present reinforcement algorithm called ResQ, algorithm designed to develop for coordinating the request from the victims and volunteers providing various assistance in Hurricane Harvey in 2017. Hurricane Harvey was a category 4 hurricane that made landfall on Texas, causing catastrophic flooding and many deaths. The model works for identifying trapped victims and rescuing volunteers and makes use of volunteers rescue strategy at its best and in time-sensitive manner. The effectiveness of this method is validated using the data set collected from social media at the time of Hurricane Harvey held at Texas in August 2017. The significance of this model is that it can schedule multiple volunteers simultaneously for the rescue

operation. This model can work with constantly changing data and achieve best performance over space and time. Alam et al. [3] apart from the analysis of textual content study about how image contents are helping in disaster response which was done. This paper presents human-labelled multimodal data set collected from Twitter at the time of seven natural disasters including wildfire, earthquakes, hurricanes and flood. Lacking labelled imaginary data is one of the issues. In this, this issue is addressed by introducing CrisisMMD, multimodal Twitter corpora consisting of lots of manually annotated tweets and images fetched at the time of seven major natural disasters like hurricanes, earthquake, wildfires and floods that happened in the year 2017 across different parts of the world. The data sets contain three types of annotations: informative versus not informative, humanitarian categories and damage severity categories. Singh et al. [2] propose a tweet classification system to support the victims in disaster. The objective of the system is to classify the high priority and low priority tweets. For these, three algorithms are used, namely (a) random forest, (b) support vector machine and (c) gradient boosting. Very poor result was shown by SVM algorithm than the other two. The classification accuracy of the system is 80%. The location of the user is evaluated from the old tweets if there is no location available in the current tweet. Markov chain method is used for finding the location. For categorizing the tweets here, only the textual contents were taken, and Internet links are not considered. The drawback here is that these Internet links will give information about some other website, in which we can gather more information about the affected area. The system identifies tweets which are related to flood in English and Hindi.

## 3 Methodology

The data set available in this work is the Chennai flood data set obtained in the year of 2015 November–December. About 40,000 tweets are taken for evaluation. Here it used a customized Twitter script for extracting backdated data by using the key word 'Chennai flood'. So many tweets are coming under this key word which is unwanted like sympathizing the situation or showing emotions. So, pre-processing of data is needed to extract tweets for requesting help and tweets for resource availability. Since it is a large data set, 20% of the data is used for training and 80% used for testing. Tweets collected have been classified using three machine learning algorithms such as linear ridge regression, SGD classifier and Naïve Bayes algorithm, and performances of three are compared with respect of accuracy, precision and recall (Fig. 1).

The methods used for study are classification and regression. Simple linear regression is an effective algorithm for the prediction of valid tweets. Linear regression is a statistical model that shows the relationship between two variables using a linear equation in which $X$ is the explanatory variable which explains the value of the other variable and it is independent. $Y$ is the dependent variable. The motive of the linear regression is to plot the straight line that is best fit for the data.

Equation of the straight line:

| | text |
|---|---|
| 4788 | This baby was lost in # chennaiflood share this photo ... http://fb.me/6YsuB3oFb |
| 6277 | Wishing you strength for today and hope for tomorrow. # ChennaiFlood http://fb.me/4U5Ts1RrE |
| 221 | Sathees_offl: RT ishafoundation: DEC 10 - Today's Medical Camps for # ChennaiRainsHelp # ChennaiFlooods vp_offl PremÃ...pic.twitter.com/rMngfapl |
| 5665 | Helicopter,Boat rescue by NDRF Tambaram 9445051077 Alandur 9445190012 Adyar 9445190013 Perungudi 9445190014 # ChennaiFlood @ selvaraghav |
| 1917 | Great tribute by @ arrahman ji for TN # ChennaiFlood # ChennaiMicro https://www.youtube.com/watch?time_continue=87&v=KwUaTVZPs4M Â... |
| 3526 | 10 # Selfless celebrities from # India united as one in aid to all the # ChennaiFlood victims! http://bit.ly/1SNONwd pic.twitter.com/RCAsshOOHo |
| 2047 | @ HyundaiIndia @ Hyundai_Global : # ChennaiFlood affected cars totaled.. Help us replacing with New car, with discount price.. Mine is 2mth old |
| 6921 | ??????? ?????????? ????? ???????? ????? ???, ??????? ??? http://eisamay.indiatimes.com/nation/flooded-chennai-traumatized-by-stray-snakes-d |
| 2917 | # BackToSchool from # ChennaiFlood # chennairain |
| 5170 | Emergency ( Medical Help : Call at 044-23452338 & 9840630903 ) # ChennaiRainHelp # ChennaiFlood # ChennaiRains |
| 3004 | Chennai floods: Madras HC seeks state report on Ã'lack of relief coordinationÃ' https://shar.es/1GgriK # ChennaiFlood # Jayafails |
| 6671 | # chennaiflood |
| 6467 | # ChennaiFlood disaster unprecedented: Venkaiah Naidu http://bit.ly/1I77tMO |
| 6314 | RT Verified- Food for 200 people available i KKNagar. Those in need please contact FESTUS - 9791183785 # ChennaiFlood # ChennaiDrowning |
| 5254 | # chennaiflood @ ArvindKejriwal JI AAPNE MLA KI BAJAY INKA VETAN BADHANE KA FESALA LIYA HOTA. INSAN K ROOP ME BHAGWANpic.twitter.com |
| 4285 | Why is Chennai Under Water http://helloexcuseme.net/chennai-flood/ # ChennaiRain # ChennaiFlood # ChennaiRainsHelppic.twitter.com/NNE0i8i |
| 6207 | # crowdsourced # ChennaiFlood map http://osm-in.github.io/flood-map/chennai.html#9.32/13.0288/80.1589/-7.9 Â... # BigData |
| 5858 | To call helicopter for help raise ur hand in Y Shape.. it will be understood by PILOT # ChennaiFlood # ChennaiRainsHelp |
| 4720 | # chennaiflood I only came to chennai with rescue team for helping flood victim . |
| 4210 | ???? ?? ??? ???? ??? ??????? ?? ??? ???? ?? # ChennaiFlood ??? ???? ?? ??????? ?? ??????? ?? ????? ???? ???? @ ISapnaMishra # helpingsouls |

**Fig. 1** Sample data set

$$Y = mx + c \tag{1}$$

so that it finds the best value for m and c so as to reduce error between actual value and predicted value. The line with least error is line of regression and the best fit line. The attributes considered in this work are set of tweets as input and set of valid tweets with respect to scope of analysis (in this case, flood) as target. Regressive analysis of data was done, and here linear ridge regression model is used for data analysis.

## *3.1 Linear Ridge Regression (LRR)*

Linear ridge regression is employed in scenarios where there is high variance of data points as compared to the data which is used to train the model. It uses the regularization technique to penalize the model. By adding a bias, it creates a new line with low variance. Thus, applying ridge regression technique, the penalty is reduced and thereby variance.

$$\min(||y - x(\theta)||_2^2 + \lambda||\theta||_2^2 \tag{2}$$

## *3.2 SGD Classifier*

Other than gradient descent, stochastic gradient descent selects some random samples for each iteration from the whole data set. The path taken to reach the minima is nosier than GD algorithm but SGD algorithm is computationally low expensive. For

performing the iteration, the sample is shuffled and randomly selected.

$$\theta_j = \theta_j - \alpha\left(\bar{y}^i - y^i\right)x_j^i \tag{3}$$

### 3.3 *Naive Bayes Classifier*

Naive Bayes is a problematic machine learning algorithm which states the conditional independency of the features in a model. This is widely used in the classification task. The theorem assumes that the features in the model are independent to each other. That is, a given change in one feature does not make any change to other feature(s) which is used in the algorithm. Naive Bayes classifier works on the basis of Bayes theorem.

Bayes theorem:

$$P\left(\frac{A}{B}\right) = P\left(\frac{B}{A}\right)P(A)/P(B) \tag{4}$$

Here $B$ is the instance, and $A$ is the hypothesis. Using this theorem, we can find the probability of $A$ happening, given that $B$ has occurred (Fig. 2).

**Finding the location of the tweet**

Tweet information filtered through machine learning model shall be passed through natural language processing algorithm to understand type of requirements/resources

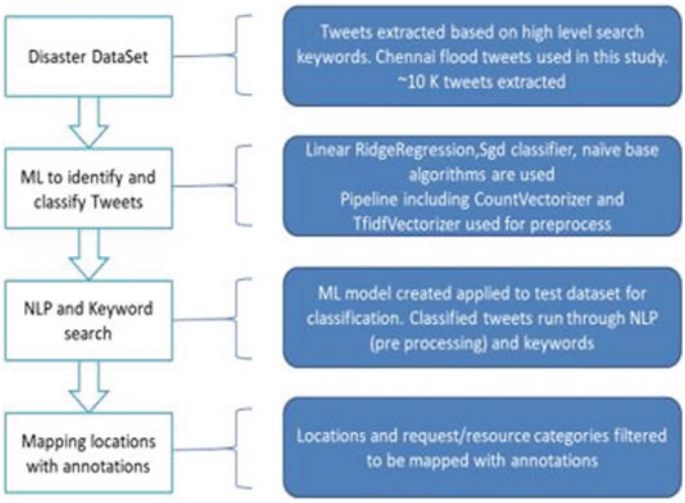



**Fig. 2** Methodology structure

and identify location in the tweet. Tweet locations from Twitter may not be always helping here as the request could be made from a different Twitter account, not the actual victim or volunteer. Also, the type and nature of request could be dependent on disaster hit area. So, these two information shall be kept as configurable so as authorities can set it as per the ground requirements.

Once analysed result is ready, geo positioning information shall be generated for the identified locations. This is achieved using module geopy, geocoders and geolocator methods.

For a given location, it shall return the latitude and longitude information. Geo locating information generated shall be then used to create maps. Mapping is done using folium module.

## 4 Implementation

Training and test data sets are converted to Pandas database for ease of processing. 20% of data set is used as training data set and is trained manually labelling tweets related to Chennai floods (true positive) as 1 and others as 0. Both trained and test data are then converted to matrix of token counts. A sparse data set is created which is then converted to dense to save memory and to improve performance. Original data will be pre-processed to further filter out random data during digitizing. Three machine learning algorithms are used to process data. Digitized train data are then fed to the machine learning model to fit the model for specific input requirements. Cross-validation methods are used to check the effectiveness of the model. Accuracy rates around 70% were achieved in the initial phase.

Tweets identified from the machine learning module are further moved through natural language processing to identify keywords in the given data set and location from where the request/response a given tweet is referring to. Depending on the nature of emergency and location, data set shall be created for filtering process. Filtering of data to a desired level is achieved with pre-processing itself.

Filtered data shall be validated against a set of preset keywords to identify nature of request and preset location to identify the exact location where assistance is required.

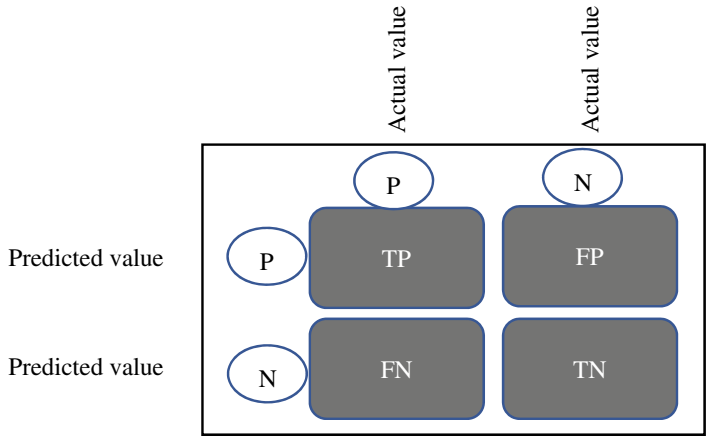
## 5 Result and Discussion

The performance of a classification model can be described using a confusion matrix. Here it used three machine learning algorithms for classification. Confusion matrix allows the visualization of the performance of an algorithm.

Evaluation matrix is defined according to the confusion matrix (Table 1).

Accuracy is the ability of the model to predict correct instance for the given observation. Precision is the fraction of the relevant instance among the retrieved

**Table 1** Confusion matrix




TP = observation and prediction positive
FN = observation positive, prediction negative
TN = observation negative, prediction is also negative
FP = observation negative, prediction positive

instance. Recall is the total amount of relevant instance that are actually retrieved (Table 2).

Validate the model with the 20% of data which kept aside. The predicted result is again analysed, and the number of true positive cases is identified. Thousand eight hundred and seven tweets got as the predicted data and in these 40 data are true positives, that means, tweets related to help offerings or tweets for help request. Fifteen data are false negative, that means, tweets related to request for help or

**Table 2** Evaluation matrix

| | |
|---|---|
| Accuracy | $(TP + TN)/(TP + TN + FP + FN)$ |
| Precision | $TP/(TP + FP)$ |
| Recall | $TP/(TP + FN)$ |

**Table 3** Confusion matrix for the three algorithms

| R. R | | SGD | | N. B | |
|---|---|---|---|---|---|
| TP | FP | TP | FP | TP | FP |
| 40 | 0 | 37 | 0 | 38 | 0 |
| TN | FN | TN | | TN | FN |
| 1719 | 15 | 1716 | 17 | 1717 | 14 |

resource availability but it is predicted as negative. This can be improved by giving more specific keywords and adding more filters in the trained data set. Thousand seven hundred and nineteen tweets are true negative means tweets which are not related to our requirement category. But that are coming with a hash tag of Chennai flood. Tweets are related to some personal statements, sympathizing situations, donations given by actors or political leaders, free performance of celebrities to raise the fund. Like this so many tweets come under the hash tag of Chennai flood or Chennai rains. So, finding the required tweets from this large amount of junk tweets is really challenging. But it is effectively done with the three machine learning algorithms. No false positive case is identified. It shows the efficiency of the model. There is a slight difference in the numbers of result by using three algorithms. It is shown under the table given (Table 4). validate the model with 20% of data so the number of data in the result is very less (Table 3).

Entire data set is divided into 20 folds for evaluation. Here performance of the system is calculated with 20-fold cross-validation. Each fold is verified using this cross-validation. The difference with respect to performance in each fold for the three algorithms is evaluated (Table 3).

The accuracy, precision and recall of the system using the three different algorithms are given below. Accuracy of the model using the three algorithm is plotted below (Figs. 3, 4, and 5). Values of the first tenfold are taken.

It is clear in the graph that accuracy is gradually increasing over folds. Above 97% accuracy is shown with linear ridge regression.

For SGD classifier and Naïve Bayes algorithm, it got only 96% accuracy. So, it is showing that linear ridge algorithm is best for this model (Figs. 6 and 7).

**Table 4** Performance comparison table (first tenfolds)

| Folds | LRR | | | SGD | | | Naïve Bayes | | |
|---|---|---|---|---|---|---|---|---|---|
| | Accuracy | Precision | Recall | Accuracy | Precision | Recall | Accuracy | Precision | Recall |
| 1 | 0.954 | 0.946 | 0.962 | 0.957 | 0.958 | 0.966 | 0.96 | 0.926 | 0.962 |
| 2 | 0.963 | 0.946 | 0.962 | 0.957 | 0.958 | 0.966 | 0.96 | 0.926 | 0.962 |
| 3 | 0.957 | 0.946 | 0.962 | 0.96 | 0.958 | 0.966 | 0.96 | 0.926 | 0.962 |
| 4 | 0.969 | 0.946 | 0.962 | 0.96 | 0.958 | 0.966 | 0.960 | 0.926 | 0.962 |
| 5 | 0.963 | 0.946 | 0.962 | 0.96 | 0.958 | 0.966 | 0.96 | 0.926 | 0.962 |
| 6 | 0.963 | 0.946 | 0.962 | 0.957 | 0.958 | 0.966 | 0.963 | 0.926 | 0.962 |
| 7 | 0.96 | 0.946 | 0.962 | 0.96 | 0.958 | 0.966 | 0.963 | 0.926 | 0.962 |
| 8 | 0.967 | 0.946 | 0.962 | 0.963 | 0.958 | 0.966 | 0.963 | 0.926 | 0.962 |
| 9 | 0.972 | 0.946 | 0.962 | 0.963 | 0.958 | 0.966 | 0.963 | 0.926 | 0.962 |
| 10 | 0.972 | 0.946 | 0.962 | 0.963 | 0.958 | 0.966 | 0.966 | 0.926 | 0.962 |

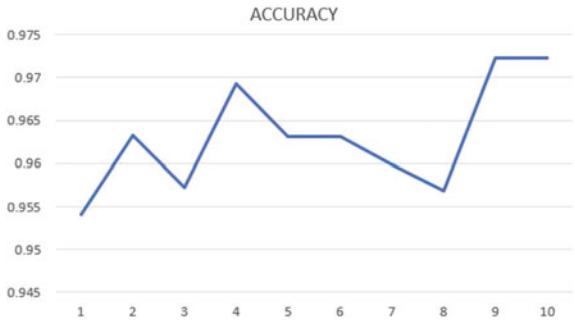



**Fig. 3** Accuracy over tenfolds using linear regression

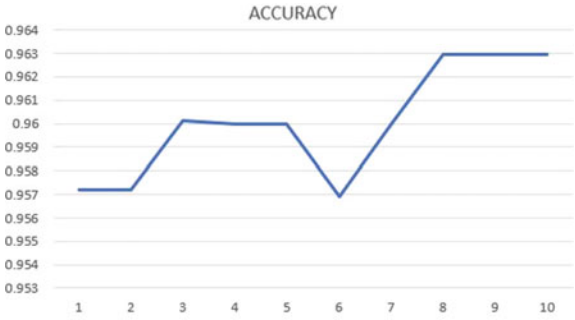



**Fig. 4** SGD classifier accuracy over tenfolds

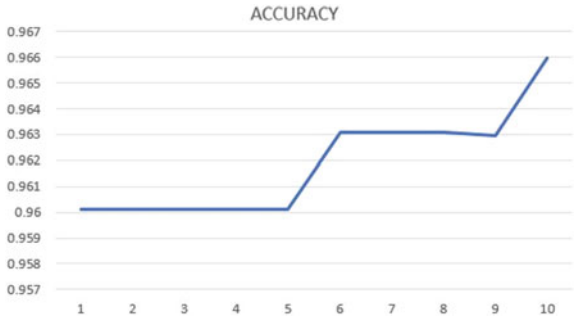



**Fig. 5** Naive Bayes algorithm accuracy over tenfolds

## 6 Conclusion

In case of emergencies, people are making use of social media platforms, and the trend is increasing as it can be observed from the recent situations faced in India. Utilizing the advancement in AI and machine learning, request in Twitter can be tracked and channelled for assistance, and the same is demonstrated in this paper. This paper proposes a model using linear ridge regression, SGD classifier and Naive Bayes algorithm to identify the valid tweets. The system is working with an accuracy of 97% accuracy for linear ridge regression and 95% for SGD classifier and Naive

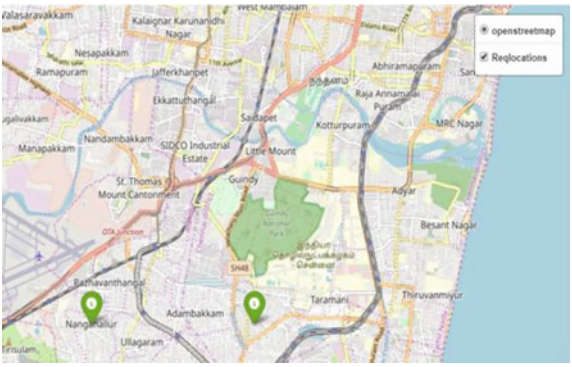



**Fig. 6** Request location

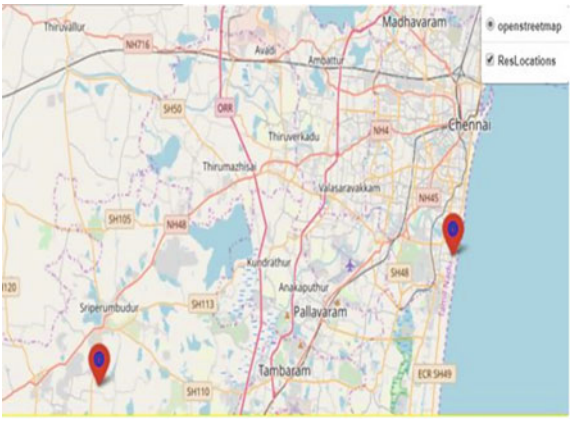



**Fig. 7** Resource location

Bayes. Anybody can easily identify the location of the help request and resource availability through this system. This system helps the efficient management of time and resources at the time of disaster. The future work is planning to map the requirement request and the resource availability according to the nearest location. The mapped result is published in other social media platform or conventional media, so that there is better access of data to the people of all walks of society. Working together with government entities for data sharing will also come in future scope.

## References

1. Abbasi MA, Kumar S, Filho JAA, Liu H (2012) Lessons learned in using social media for disaster relief—ASU crisis response game. In: Yang SJ, Greenberg AM, Endsley M (eds) Social computing, behavioral—cultural modeling and prediction. SBP 2012. Lecture notes in computer science, vol 7227. Springer, Berlin, Heidelberg

2. Nguyen LH, Yang Z, Zhu J, Li J, Jin F Department of Computer Science, Texas Tech University; George Washington University, Department of Civil, Environmental and Construction Engineering, Texas Tech University
3. Alam F, Ofli F, Imran M Qatar Computing Research Institute, HBKU, Doha, Qatar
4. Castillo C, Mendoza M, Poblete B (2011) Information credibility on Twitter. Paper presented at the www2011, Hyderabad, India
5. Singh JP, Dwivedi YK, Rana NP Event classification and location prediction from tweets during disasters
6. Puterman ML (1994) Markov decision processes: discrete stochastic dynamic programming, 1st edn. Wiley, New York, NY, USA
7. McMinn AJ, Tsvetkov D, Yordanov T, Patterson A, Szk R, Rodriguez Perez JA, Jose JM (2014) An interactive interface for visualizing events on Twitter. In: Proceedings of the 37th international ACM SIGIR conference on research & development in information retrieval, Gold Coast, Australia, 2014, pp 1271–1272
8. Petrović S, Osborne M, Lavrenko V (2012) Using paraphrases for improving first story detection in news and Twitter. In: Proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, Montreal, Canada, 2012, pp 338–346

**V. G. Dhanya** is pursing Masters programme in Computer Science, in KCG College of Technologies, Anna University, and presented a paper based on machine learning in the National Conference NCAI 2020. Areas of interest are data analysis and machine learning.

**Minu Susan Jacob** is an assistant professor in the KCG College of Technology and is pursuing research in data analytics and machine learning.

# Sentimental Analysis on Twitter Data of Political Domain


**Seenaiah Pedipina, S. Sankar, and R. Dhanalakshmi**

**Abstract** The arrival of social media has initiated the platform for the public to express their views and to share their emotions. In addition to this, smart phones and mobile communication technologies have emerged and become as a tool to spend more time than earlier on social media to be in touch with their friends. The way of expressing the thoughts, attitude, and feelings is changed drastically due to that short messages and emoticons utilization has been increased through social media. There have been no constraints in using the micro blogging services like Facebook, Twitter, etc. Thus, these messages are the views of people used to describe their behavior. Nowadays political party workers and leaders also started spending more time on Twitter, Facebook, and blogs to be in touch with the public. Hence, political parties or social media campaigners started analyzing for the solutions to utilize the public opinions or views about their own political party. A program is developed which collects the reviews or text posts of people and transfer it to the next module called pre-processing or cleansing. This module is to remove URLs, special symbols, Junk text, stop words, and tokenize the sentences and to perform the stemming process. Word2vec model is used to perform the word embedding to convert the text data into numerical vector format to transfer it to Recurrent Neural Network. The sigmoid function is used as an activation function to classify and polarity of positive and negative sentiments have been evaluated. The effectiveness of the proposed system has been proved through experimental analysis in this paper. This proposed system will be used to understand the winning chances of political party and to analyze the response of the public on particular political decision during election campaign.

**Keywords** Sentiment mining · Sentiment analysis · Word2vec · Artificial neural network · Classification

S. Pedipina (✉) · S. Sankar · R. Dhanalakshmi
KCG College of Technology, Chennai, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_17

# 1 Introduction

The utilization of the internet has got increased rapidly and social media has a major contribution to users. Social media has created the platform for the users to share their feelings, status, and emotions; this includes feedback about a particular product or movie, experience of a particular service, opinion on the government reaction over social issues, etc. Users post their opinions through microblogging sites or social media as their expression. As the usage of social media is high, hence it generates a huge amount of data, which can be used to identify the collective sentiment of people. Thus, it has become the best option for taking important decisions related to their business, to run the political campaign according to the people's interest, and for good product consumption.

Earlier political parties used to conduct surveys collecting sample opinions in person to understand public satisfaction on government but the consumption of time and cost is high in this approach. In these days, even politicians are creating their official pages on the social media especially on Twitter to stay in touch with the people. Public also responds in social media over the government performance, social welfare schemes implementation and public sector services. Thus, political parties started leveraging the social media platforms by extracting the data from the micro blogging services, which is easy and more effective way. Data also trustworthy to analyze as the social media users share their genuine feelings, feedbacks, and emotions. In addition, there are many effective tools available to collect, the data from these sites, which provides by same sites or third-party vendors.

As everyone knows that Twitter provides the access to the raw format of the tweets text data, it is essentials for the analyzer to evaluate and analyze the data accurately. Data is collected from different sources, chances are more to have junk or irrelevant data [1]. Thus, to remove this effective data cleansing and pre-processing should be applied for removing bad data. Sometimes data would be good but it does not help us in classifying the sentiment polarity. More the data is clean more the accuracy would come from machine learning algorithms. Hence, there is very much need in design to have effective data cleansing techniques while pre-processing the data.

## *1.1 Sentiment Analysis*

Sentimental analysis has become the buzzword in the natural language processing as the utilization of this analysis has got spread widely to different domains. This is a process to evaluate the text data and extract the opinion polarity out of the text, and hence this process is called opinion mining. There are three main opinion polarities called positive, negative, and neutral that can be extracted from the text user interested in. This process is used to understand the opinion of the writer or speaker with respect to the subject text [2]. Sentiment analysis is categorized into three levels depends on the text data to which sentiment should be matched.

If the sentiment is assigned to a document, evaluating the whole text belongs to that document then it is called sentiment analysis at document level. For example, if the data is extracted from an internet blog and if the the user needs to identify the polarity of the whole text then this type of sentiment analysis can be used.

If the sentiment is assigned to each sentence in multi-sentenced document then it is called sentiment analysis in sentence level. For example when user wants to assign a opinion polarity for each tweet extracted from twitter belongs to a particular domain (political domain in this paper) then this type of analysis can be used.

If the sentiment is assigned to each phrase in a text then it is called phrase level or aspect level sentiment analysis. In this level, every word is evaluated and assigned the sentiment. When user wants to count the number of positive or negative words present in text then this process can be used.

There are legacy and modern techniques available to perform sentiment analysis. The lexicon-based approach is the legacy technique in which scores are assigned to each lexicon of words or terms. Further, the sentiments can be assigned to text or words using the pre-assigned scores. In lexicon-based, we have dictionary-based and corpus-based techniques. In dictionary-based approach, sentiment is assigned by identifying the presence of signaling sentiment words. The corpus based is data-driven approach where sentiments are labeled based on the contexts.

In this proposed system, a recurrent neural network model is used along with Long Shor Term Memory (LSTM) layer to train and classify the political tweets data with positive, neutral, and negative polarities.

## 2 Review of Literature

The sentimental analysis has got the huge attention from different domains like marketing, entertainment, digital and political which are actively related to social media messages. IT department has become essential for every company or party to publish messages to their followers and analyze the response. Hence Sentimental analysis is playing a very important role in helping the sources to understand the public response on particular products of a brand or tweets that are posted by a political leader. The different conclusions or insights that companies can understand in this are a mobile company can predict the demand for their upcoming mobile by understanding the positive comments posted on Facebook or Twitter for their posts or advertisements. Audience can understand the product reviews by analyzing the sentiments behind the reviews that are posted for that particular product or a political leader can understand how well the public is satisfied on the new government policy by collecting and predicting the sentiments of their tweets. As such there are many problem statements in which sentimental analysis is helping in different sectors. Hence, there is a lot of scope to research and identify the better solutions than the existing to classify and predict the sentiments. There are different methods and lexical based methods to analyze sentiments from tweets that are in natural language format in order to identify the polarity: positive, neutral, and negative and it. The

lexicon-based method [4], refers two dictionaries; one for bag of words and other for emoticons.

In dictionary approach, list of words has to be built manually and it is a time-consuming process and also requires frequent alterations. The disadvantage of this model is that it cannot assign the polarity for the word which is out of the pre-defined dictionary which leads to bad accuracy of the model. This makes the programmers look for different approaches that are using machine learning. Comparing the accuracy of sentiment analysis using this traditional approach and supervised machine learning approach [5] of Naive Bayes for assigning the positive and negative polarities to tweets data. Existing researches have proved that machine-learning approach is better accurate than lexicon-based method. However, there are disadvantages in machine learning approach as well. The accuracy of naive Baye's model is always proportional to the dictionary size, so the word count should always be high [6] to increase the model accuracy. The other mostly used algorithm in text classification is support vector machines which gives more accuracy than Naïve Bayes but developing SVM model is not easy compared to Naïve Baye's and this requires large amount of training data, at the same time testing the model is also slow. This model also finds difficulty in identifying important words for classification [5]. In natural language processing, each input item may contain multiple sentences, which gives the exact meaning only if they are processed together. So maintaining the context between these kinds of sentences is very much important to get the better accuracy. Traditional machine learning algorithms that are support vector machines, Naïve Bayes are failing to maintain the meaning when they face these data items in input. To get rid of these problems, in this proposed system, recurrent neural network deep learning algorithm is used.

The text data has to be transformed into numerical vector format as input to deep neural network, to accomplish this there are several word embedding techniques. One-hot encoder approach is one of them used mostly for vector representation. In this approach binary values 0 or 1 are assigned for all the elements. 1 is assigned to the index of the integer and remaining all the values would be 0. However, drawback of this method is that context cannot be maintained between the words present in the same sentence. Word2vec is the most widely used word embedding technique and so it is used in this proposed system as well to convert the text data into numerical vector with the dimension of 200.

The following existing experiments have been evaluated to select the effective techniques in pre-processing, word embedding, deep neural network model in different modules in the proposed system.

Kumar et al., proposed a machine learning approach to analyze customer satisfaction from airline tweets [7] which is about investigation of twitter data collected for airline business. Some regular main airline corporations have been chosen across the world based on their number of tweets and followers on the Twitter. Twitter data has been collected, cleansed and transformed into numerical vector format using *n*-gram approach for further analysis. Initially, numerous artificial neural network models were applied and performed testing along with support vector machines on hybrid word embedding, trained and pre-trained datasets. Artificial neural network4 model

has given more accuracy compared to all other developed SVM and ANN architectures models. Further, SVM and ANN4 have been also analyzed using different *n*-gram representation of Twitter data. It has been identified that *n*-gram model with $n = 4$, gives better accuracy for both SVM and ANN4 models. At the end, the convolution neural network model has been developed to predict the sentiments polarity behind the tweets data even though CNN models are mostly used for image classification and it provides a drastic development in sentiment prediction of text classification model performance.

Chen et al., has proposed a system for classifying the sentiments using intensive sentiment supplement information [8] and negative meaning data. This system has addressed the semantic composition problem of current deep neural network related models in sentiment classification. Particularly, this is designed for the major impacts of intensive and negative words through a GRU or LSTM network, that are modeling negative expression and intensive expression. The effect to the sentiment reversing on succeeding expression due to the negative words like "not" and "never" is analyzed in this system by using backward LSTM model. These models are integrated into three neural networks as LSTM, CNN, and Char SCNN to evaluate the efficiency of the above operations, and denoted these new models as NIM-LSTM, NIM-CNN, and NIM-Char SCNN, where "NIM" means "Negative and Intensive Modeling."

Namugera et al. [9] proposed a system to mine the text data and determining the sentiments. This system explores the social media sites like Twitter and Facebook usage by Uganda media houses. This system has examined the tweets data belongs to print and digital media companies in Uganda from Twitter to form the sensible insights using text data mining approaches. The study had proved that tweets with negative sentiment are highly engaged with users. This study has also analyzed the rate at which retweet count or mentions has increased for tweets with negative and positive sentiments and proved that tweets with negative polarity spread faster than the positive tweets. Thus, there is very much need to monitor these negative spreads closely to control the fake news propagation. This study has also recommended to extend the proposed model using beg data analytics for sentiment prediction.

After the study, it is understood that in the social media domain, sentiment analysis has more scope to research in this area. There has been already sentiment analysis is done on the user reviews related to political domain with respect to the government decisions and political issues. As we got the ability to collect and label the huge amount of data, we can leverage the deep neural network models that are recurrent neural network and convolutional neural network. Convolutions networks are best suited for image data analysis. Hence RNN has been selected in this proposed system in addition LSTM layer also provides the capability of maintaining the context between sentences in order to train and predict the sentiment polarity more accurately.

## 3 System Architecture

This section explains the different modules available in the proposed approach. The proposed system architecture is shown in Fig. 1. This approach has designed to extract the messages posted by users related to political parties, government policies, and political diplomats from the social network sites like Facebook or Twitter and then perform pre-processing by removing the unwanted and junk data, which does not carry any information related to opinion polarity along with the tweets from extracted data. This process is also called cleansing. After performing cleansing, Recurrent Neural Network with LSTM model is analyzed and then the sentiment behind each tweet is predicted. There major steps are followed during the architecture implementation:

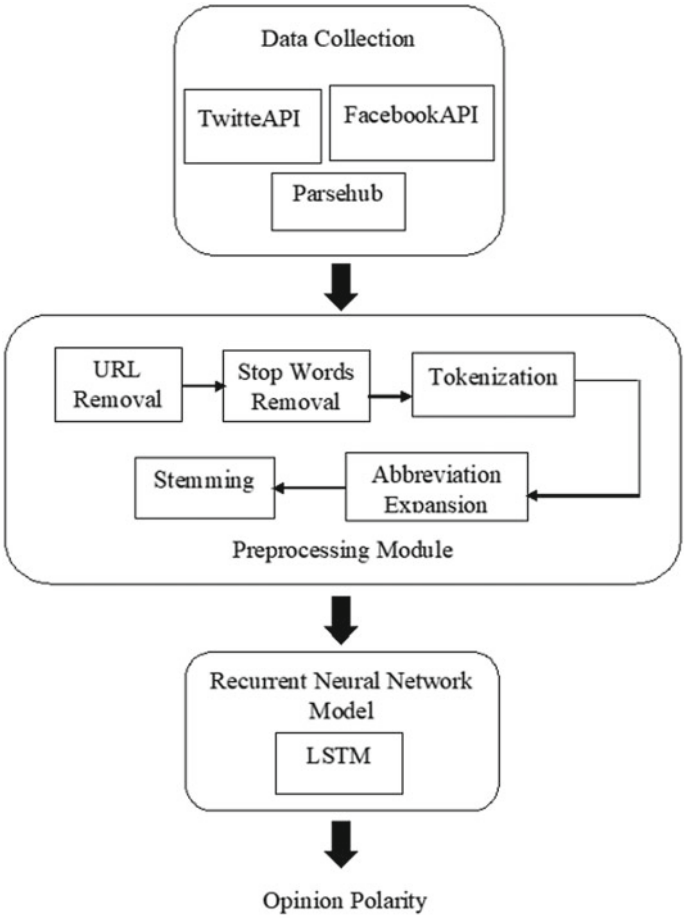



**Fig. 1** Proposed system architecture

1. Data Collection using Python API
2. Preprocessing
3. Classify each tweet as positive, negative or neutral using RNN with LSTM.

## *3.1 Data Collection*

The very first step in the architecture is that collecting the twitter data that is posted by public related to political domain. There are several tools and APIs available in order to study the sentiments of the user. Web scraping or web crawling is the technique for extracting the data from web blogs, feeds. Data has been extracted from the widely used social media web sites like Facebook and Twitter. Tweepy OAuth API has been used to import the Twitter data. Facebook's Graph API is used to gather the user posts from Facebook. Data is also collected from public domains sites and different blogs using web crawling API called ScraPy using Python.

## *3.2 Preprocessing*

The data that is collected in the Data Collection step is not structured or semistructured as it contains unusable parts such as web URLs, advertisements, HTML tags, stop words, scripts, and junk characters. The processes of removing all these uninformative parts for sentiment analysis are called pre-processing. In this process, data is cleaned and passed as input to the upcoming steps in the system. The data dimensionality plays important role in classification process, words that are presented in the data is represented as one dimension. The dimensionality also will be increased due to inappropriate words. Hence, the preprocessing or cleansing process is very much essential to get better accuracy out of any Neural Network model.

There are different Python libraries available to perform cleansing process. NLTK is the most used Python library in pre-processing the data. Python regular expressions (re library) can also be used by passing the text patterns to replace the bad data.

Regular expression package is used to delete special characters, URLs and other junk data. NLTK library is used to perform Tokenization, Stop word removal, abbreviation expansion and stemming. The complete preprocessing procedure contains numerous steps: Removal of URL, white space and special characters, removing other words which doesn't help in sentiment classification, tokenization, and parts of speech tagging.

The sequence of steps given as below in preprocessing:

*Step 1: Removing junk characters and URLs*
*Step 2: Removing special characters and numbers*
*Step 3: Removing stop words from text*
*Step 4: Tokenizing the tweets into sequence of words.*

### 3.3 Word Embedding

Usually, neural network models accept the input in numerical vector format. Word embedding is the process of converting the text data into multi-dimensional numerical vector format. In this process, 200 dimension input vector is used. After performing preprocessing to remove all the unrelated data from the collected data, this text data needs to be transformed into numerical format in order to pass it to the recurrent neural network for the sentiment polarity mapping process. There are many approaches to perform word embedding. One-hot encoding is the legacy approach which is widely used method for performing word In this method the number of words equals the array of size in the document is preserved. Every word has been mapped to 0 or 1 is characterized by transforming bit from 0 to 1 in the position of that word in the array. The disadvantage of this model is that it cannot maintain the context between multiple sentences when there is dependency between them. The context should be maintained between all the words to understand the correct meaning lying in the text, while converting the text into numerical vector format. The word2vec model is used to avoid this problem in word embedding. This is a shallow and two-layer neural network model designed and mostly contributed byGoogle. These models can be trained to reconstruct the linguistic contexts and meaning of words in between the sentences [10].

Word2vec model can be trained using dataset of textual tweets and posts that are extracted from Twitter and Facebook. Each word in text message is characterized to numerical vector with the dimension of 200. These vectors are further passed to the proposed model recurrent neural network as input. The input tokenized sentences are further converted to order of vectors from order of words with the dimension of 200.

$$-\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{-c \le j \le c \\ f \ne 0}} \log p\left(w_{t+j}|w_t\right) \tag{1}$$

There are different techniques to perform word embedding in word2vec itself. Here Skip-gram model is used in this proposed model. The word's probability and its context can be increased by using skip-gram model. In Eq. (1), '$w$' represents the centered (at $t$) word and '$c$' represents the training context size. The high number of vectors also affects in training phase where there are more words, which does not help in assigning the sentiment polarity. To avoid this, TF-IDF technique is used to reduce the irrelevant vectors while maintaining the each word weightage. This method also evaluates the average weight of the vectors of all the words. Hence, every input sequence is now transformed to vector representation with the dimension of 200 along with the meaning within the sequence.

### 3.4 *Recurrent Neural Network (ANN)*

Recurrent Neural Networks (RNN) is multi-layer fully connected deep neural networks. This model is used for text classification process. This model has a capacity to evaluate the relation between input data and target variables to identify the patterns associated. There are different factors that plays important role in behavior of the recurrent neural network. They are input data, weights, loss and activation functions. Sigmoid activation function has been used in this model to describe the output of every neuron [11]. Once after text data is represented by the numerical vectors format; these numerical vectors are later passed to feed the Recurrent Neural Network model in training phase. RNN leverages the Long Short Term Memory (LSTM) layer to store the data in memory when there is dependency between multiple sentences to maintain the context of the text while processing is in progress.

RNN contains of 200 inputs, since every sentence is represented using 200 vectors. These inputs are then moved to hidden layer, which is consisting of 100 neurons. Final result is evaluated by the activation function (sigmoid function in this case), which gives output in the form of 0 or 1 (Fig. 2).

The sequence of steps given as below in Artificial Neural Network phase:

1. *Input: Numerical vector of size 200*
2. *Output: Sentiment Polarity*
3. *Initializing weights randomly to the inputs,*

*Input layer evaluates total input by*

$$I = \sum_{i=1}^{n} W_i X_i \tag{2}$$

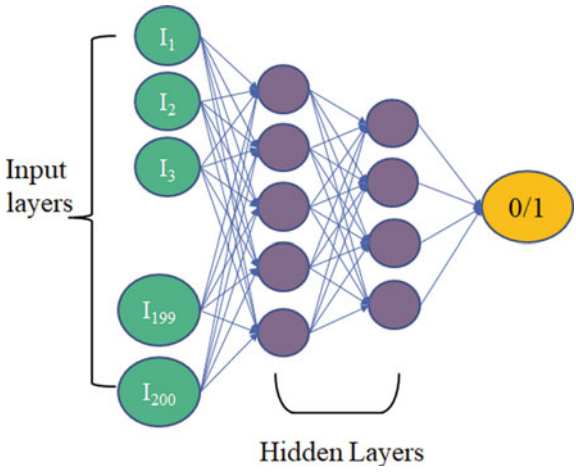



**Fig. 2** Recurrent neural network with 200 inputs, 2 hidden layers, and 1 output and single neuron with activation function

4. *Hidden layers perform processing & backward propagation algorithm attempts to decrease the error, the error function is given below*

$$E = \frac{1}{2}\left(\sum_{k=1}^{m}(y_k - d_k)^2\right) \tag{3}$$

5. *Sigmoid function is used as an activation function to produce the output*

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

## 4 Results and Discussion

This phase is about discussing the results of the proposed model. Usually, Recurrent Neural Network is suitable for text processing for example predictions and word generation. And also prior studies prove that RNN performs well on text classification.

To train the model a dataset that has 4400 tweets of Indonesia politics has been downloaded and performed pre-processing, word embedding, and passed the numerical vector data to the model. Model has given the accuracy of 85%.

In order to predict the sentiments of the users for the specific political party, around 2350 were collected from Twitter related to Indian Prime Minister Narendra Modi. The collected data is Pre-processed, transformed into numerical vector, and then passed to the trained recurrent neural network model to perform the prediction. After predicting the underlying sentiments positive, neutral, and negative, the count of each sentiment has been represented in Fig. 3. The Sigmoid activation function has given the output with the score between 0 and 1. The tweet that has the prediction

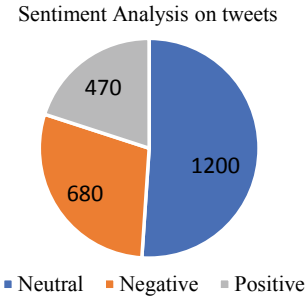



**Fig. 3** Count of positive, negative, and neutral tweets

score close to 0 is categorized as Negative sentiment, tweet that has the score close to 1 is categorized as positive sentiment and to classify the neutral sentiment, the scores for few tweets have been evaluated in the middle range and classified them as neutral.

## 5 Conclusion and Future Scope

Social media has become the platform for public to share views and opinions genuinely. So, political parties can consider it as feedback of the users for their decisions with respect to the government or their own political party. Extracting and evaluating the text to understand the sentiments always help them to take better decisions in attracting the public. The major social media platforms like Facebook, Twitter are providing the access to the users data and tools to collect the data for analysis. The existing proposed systems have been classifying the sentiments by using legacy approaches that are support vector machines, naïve Bayes machine learning algorithms. This paper has extended the analysis using deep learning models like recurrent neural network and LSTM neural network models. The same data has been used for all the above-mentioned models and classified the opinion polarity and all the performance metrics have been analyzed. This proposed model with LSTM layer has outperformed all the legacy algorithms. Different word embedding techniques one hot encoding and word2vec models are used to convert the text data into numerical vectors to pass as input to the model and model has given better accuracy with Word2vec embedding technique. In the future, better data cleansing techniques can be used. Increasing the hidden layers and data size also can increase the accuracy of the model. There are different state of art word embedding techniques like Glove and BERT available to try in the convolutional neural network to leverage the best of deep learning techniques to train and predict the sentiments. Once the model is ready with highest possible accuracy then this helps the political parties to perform sentiment analysis in real time to plan election campaigning to increase the winning chances in the elections.

## References

1. Hernandez-Suarez A et al (2017) Predicting political mood tendencies based on Twitter data. In: 2017 5th international workshop on biometrics and forensics (IWBF), Coventry, 2017, pp 1–6
2. Schouten K, van der Weijde O, Frasincar F, Dekker R (2018) Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. IEEE Trans Cybern 48(4):1263–1275
3. Devika MD, Sunitha C, Ganesh A (2016) Sentiment analysis: a comparative study on different approaches. Procedia Comput Sci 87:44–49. ISSN 1877-0509

4. Walha A, Ghozzi F, Gargouri F (2016) A Lexicon approach to multidimensional analysis of tweets opinion. In: 2016 IEEE/ACS 13th international conference of computer systems and applications (AICCSA), Agadir, 2016, pp 1–8
5. Bhuta S, Doshi A, Doshi U, Narvekar M (2014) A review of techniques for sentiment analysis of Twitter data. In: 2014 international conference on issues and challenges in intelligent computing techniques (ICICT), Ghaziabad, 2014, pp 583–591
6. Islam T, Bappy AR, Rahman T, Uddin MS (2016) Filtering political sentiment in social media from textual information. In: 2016 5th international conference on informatics, electronics and vision (ICIEV), Dhaka, 2016, pp 663–666
7. Kumar S, Zymbler M (2019) A machine learning approach to analyze customer satisfaction from airline tweets. J Big Data 6:62
8. Chen X, Rao Y, Xie H, Wang FL, Zhao Y, Yin J (2019) Sentiment classification using negative and intensive sentiment supplement information. Data Sci Eng. https://doi.org/10.1007/s41019-019-0094-8
9. Namugera F, Wesonga R, Jehopio P (2019) Text mining and determinants of sentiments: Twitter social media usage by traditional media houses in Uganda. Comput Soc Netw 6:3
10. Mikolov T, Chen K, Corrado GS, Dean J (2013). Efficient estimation of word representations in vector space. In: Proceedings of workshop at ICLR
11. Nielsen M (2015) Neural networks and deep learning. Determination Press, San Francisco, CA

# Cloud-Based Smart Environment Using Internet of Things (IoT)


**E. Laxmi Lydia, Jose Moses Gummadi, Sharmili Nukapeyi, Sumalatha Lingamgunta, A. Krishna Mohan, and Ravuri Daniel**

**Abstract** Internet of things (IoT) is a primary computational paradigm to develop a smart environment in every area of health, city, factory, and home in our daily lives. It incorporates wireless transmissions to all sensor devices through the internet. Equipping a smart environment to the society, IoT as the primary source provides alternative diversified communicating characteristics. Its ecosystem is the solution to all communication technologies as well as designed architectures. This paper deals with distinct core requirements to generate reusable features and technologies to develop a smart environment. Technological architectures like Radio Frequency Identification (RFID) and Constrained Node Network (CNN) are identified to enhance the Internet of things. This paper also describes the necessity of having smart environment sensors with the Internet of Things (IoT). This shows the involvement of a smart environment crossing all communicative disputes from the technical and informative perspective that desires to fulfill the efforts of the people in the coming years.

**Keywords** IoT · Smart environment · Cloud computing · RFID · CNN

E. Laxmi Lydia (✉)
Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India

J. M. Gummadi
Department of CSE, VFSTR (Deemed to be University), Guntur, India

S. Nukapeyi
Department of Computer Science Engineering, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

S. Lingamgunta · A. Krishna Mohan
Department of Computer Science & Engineering, UCEK (A), JNTUK-Kakinada, Kakinada, India

Department of Computer Science and Engineering, JNTUK, Kakinada, Andhra Pradesh, India

S. Lingamgunta
e-mail: sumalatha@jntucek.ac.in

R. Daniel
Department of Computer Science and Engineering, Bapatla Engineering College (Autonomous), Bapatla, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_18

## 1 Introduction

The measures of IoT computations and prototypes for communication transmission on connecting frequent objects to the internet are sustained by the establishment of resource-constrain tools. These incorporate and facilitate intelligent systems using sensors and actuators. Intelligent systems gather physical world information to perform actions subsequently. Most of the efficient advantages of IoT have resource management, upgraded productivity, and elevated quality of life for the smart environment of the large human population. The beginning of a smart environment starts with enhancing smart homes, cities, factories, and surrounding areas that relate to our daily activities. IoT tools make possible multiple transmissions of data through Radio Frequency Identification and Constrained node network technologies. These technologies have various features and plenty of heterogeneous solutions for IoT communications. Smart environments gather information from different sources based on different scales with disparate prerequisites. Therefore, information is retrieved depending upon unique domain connectives.

IoT appliances demand maintenance and assist the sensors to handle data for storing, managing, and processing. The various data flow through consistent approaches like cloud computing and Fog computing exist.

The influence of the internet on remote areas comprises scalable solutions with low latency for more number of devices are connected uses fog computing. Real-time interactions for smart monitoring intend to have smart homes through cloud computing. This allows us to control complementary data processing and storage operations but not absolute to the latency such as operating television, turning on the bulb, AC so on through remote control. For explicit interaction of devices, it is mandatory to have either cloud computing or fog computing. Smart cities were operated through cloud and fog techniques. Low latency helps us to have more number of IoT tools with potential information within the city. Production of smart factories invokes advanced sensors and actuators for the internet connectivity among large administration systems and companies through gateway using stream of data. Smart factory applications like Enterprise Resource planning, manufacturing execution systems develop smart products by monitoring machine-health parameters.

To aid all smart applicational manufacturing products, computing analytics help to process data in real-time with capability apprehension. The waiting time of the data processing is the primary concept to progress manufacturing applications as it ranges within milliseconds. Instead of continuous execution of all jobs in cloud computing, smart technologies make use of fog computing to offload some of the data processing jobs. This results in introducing low latency and low operational investment i.e., a smart manufacturing company. Industry 4.0 realize to have cloud computing to share the data from the internet to all enterprise and suppliers which hold manufacturing process using cloud services. The major intention of the expanded transparency through the supply chain and procedural networks lead to the automated production environment.

## 2 Literature Survey

The dominance to empower users in smart cities reduces the deployment rate and enhances the development of the capillary sensory apparatus. Nevertheless, the beginning of the smart cities uses these mechanisms uniformly to progress and test at many distinct exploratory platforms. Along with the frequent mechanisms the emerging edge computing allows different forms of network organization to sense the environment [2].

Some of the political issues in the cities obstruct the formation of the smart city due to an assortment of technical problems and lack of technologies [1]. The interoperability of high-tech solutions in the area of the Internet of Things extensively identifies the basic technological promoter of smart cities. The modernized progression of evolution uses mobile applications to sense models associate with valuable functionality to sense information from the cities for intimate smartphones [3].

Smart homes provide smart services that vary extensively. Assistive services and management services are the services that help to meet the needs of user mostly elderly people, disabled. These satisfy to furnish the e-health services [4]. The frequent explosion of technological modifications has led to the evolution of smart sensors, advanced computing approaches (cloud computing, big data analytics), internet of things possibly correlated to smart home healthcare projects. These smart home-based projects were carried out in research and commercial systems using two primary real-time telemonitoring methods like video examinations/consultations and also diverse bio-signals [7].

Collection of information from many systems as input, for example, computer vision frame smart environments. The obligation of the entire process acknowledges the smart environment from a human perspective directly or indirectly [5]. According to ICT, mostly contradict to all fields and challenges, and point schemes of smart cities, some precisely extend to smart energy [12], e-government [11], e-health [6],e-culture, smart mobility, e-tourism [16].

Electro-Mechanical Systems (MEMS) use inertial sensors as the most unique measures for better human observations. Depth cameras give more intelligence and evaluate more efficiently [7].

The moderate potential wireless networking is to describe instability as a concern of hasty improvement of ingenious IoT resolutions that unfold the restricted modern technologies. The eventual consideration and the work across the major improvement in limited issues. IoT technologies and systems acknowledge the exchange of data to initiate a smart environment astounded by artificial and enveloping intelligence [8].

For the entire population, quality-of-life is increased through the assistance of adequate resource management, complemented productivity IoT [9]. Later on, the applications are encouraged to enterprise the precision of endemic diseases and deterioration [10]. This can enhance the living standard of smart cities and their reparations.

Intelligent systems [13] for the smart environment include smart remote driving technologies for the benefit of the materialized business standard of Industry 4.0.

Smart homes, manufactured systems, working offices acquire the assistance of common application with the process of acquiring data and dealing with disparate compatible context information.

One of the most complex scheme to develop smart cities and smart environment come across many dominions [14, 15] that incorporate the economy, energy, mobility, governance of a huge number of correlated disputes and implicate collective actors such as city service providers, administrators, operators, citizens with conceivable challenging objectives.

## 3 Methodology

### *3.1 RFID Technology*

Radio Frequency Identification is a noticeable technology that implements a selection of the immense range of frequencies, merge with multiple device types and communication protocols. It is a standardized model to large international organizations like ISO (International Organization for Standardization), ITU (International Telecommunication Union), IEC (International Electrotechnical Commission), and also national wide organizations like DIN (Deutsches Institut for Normung), JIS (Japanese Industrial Standards) and SINIAV (National Vehicle Identification System). A trivial method is been accepted by all technologies that are equivalent to the Frequency band.

RFID technologies have a low-frequency band with 125–134.5 kHz, a high-frequency band with 13.553–13.567 MHz, and Ultra-high frequency band with 433 MHz and 858–960 MHz, Microwave with 2.4–2.454 GHz 5.725–50,875 GHz. RFID technologies have Passive, semi-Passive, Active, Sensor Tag chip type for transmission, signal modulations, and Chipless time-domain frequency. RFID technologies that are used by the IoT are ISO 14443 (13.56 MHz), ISO 18000-63 (858–960 MHz), ISO 18000-7 (433 MHz), EN 300 220-2007 for different applications such as personal identification, Online Payments, ticket booking, Access control, and security, tracking logistics, retail/consumer applications, real-time location tracking.

Most of the IoT applications utilize any three of the shared types of RFID technologies that match the number of collaborating devices at the same location with supported speed by preventing collisions.

- **Radio Frequency Identification (Identifier schemes)**

The expansion of the open IoT systems consolidate any number of stakeholders with the scalable operation. These RFID identifiers retrieve tags by interpreting them in a general way. Use of widespread standard identifier schemes requests to have RFID among the mixture of material objects, digital artifacts, and their locations with a feasible point of view. SG-1 has matured Electronic Product codes globally through object identifiers at ISO/ITU Organizations. Identifiers from Japan use Ubiquitous

IDs to employ RFID technology. Industries classify tagged billions of objects using barcode encodings. Patterns and codes of the individual scheme may not have the same for all identifiers. EPC system follows a specified prefix code for all its users i.e., 00110000. The remainder of the code generates 96 bit length as Serialized Global Trade Item Number which is always ordered in a stratified process that enables the code within the organizations with allotted authorization.

- **Radio Frequency Identification (Identifier resolution systems)**

Internet of Things (IoT) system integrates with RFID. It tags the existence of individual entities to analyze the information related to the identifier, privacy protection, and secured operations provided by the control access. This capacity and the effectiveness of the system is indicated as Identifier Resolution Service. The ongoing scheme specifications for services use Identifier Resolution Service. Contradiction to the IoT applications, they need to be open and comprehensive.

## *3.2 Constrained Node Network (CNN) Technology*

Technologies like RFID refer to communicate with other objects, but CNN is preferably used to communicate with the physical world. These are the sensing computer tiny devices with a max 16-bit processor and ~10 KB RAM with some computational constraints, energy constraints (for example batteries with limited energy source). The communication transmission can be done through wired or wireless technologies. CNN technology mostly works under MAC and physical layer functionalities. Every innovation may have explicit qualities and might be more qualified for a restricted arrangement of situations. The arrangement of technological advancement consists of

- ***IEEE 802.15.4*** is a wireless technology meant to allow tracking and manipulating various applications for WPAN. Its first publication point to standardized low-rate transmissions and low energy consumption. It was not produced to work for only a particular domain, rather preferred as universal technology with significant protocol architectures by upholding Ipv6 and protocol architecture with non-IP-based protocol architectures. Smart environments with optimized technology like IEEE 802.15.4 were designed to neglect destructions in Industrial environments. This mode of implementation is referred to as Time slotted channel hopping a specified protocol.
- ***BLE*** was launched in 2010 with a very low-electricity variant of conventional Bluetooth. About the partial use of hardware, a tool that is supported by the conventional Bluetooth also supports the low additional price of BLE. Consequently, this will leverage its large existence in smartphones, possibly accumulated information can be sent to nearby sensing devices and actuators. Smartphones that interact using tools, actuators over the Internet have a gateway. This can be further designed for most of electronic devices, wearable.

- ***ITU-T G.9959*** is a wireless Z-wave technology that clarifies the stack protocol mostly innovated to model smart homes i.e., home automation.
- ***DECT-ULE*** is a technology meant for voice and data transmissions through the gateway. It generates low-energy and enables communications for indoor cordless telephony connected to all sensors, actuators by manipulating the robust latency of DECT appliances in the home.
- ***NFC*** is a wireless technology that implements reader mode, peer to peer communications and satisfies multiple communication modes of data communication within a short range of ~10 cm. The intrinsic security properties of NFC provide maximum rejection of unauthorized devices that intend to grab transmitted data.
- ***IEEE 802.11*** is referred to as WLAN wireless communication, designed to have various power-saving techniques with low power consumption, energy-constrained tools. It also enables multiple compilations of sensor devices for data collection namely smart grid.
- ***LoRaWAN*** is a wireless technology referred to LPWAN. Its communication through the physical layer transmits data within 10s of km. It follows star topology as it transfers data between hundreds of communicating devices with low-cost infrastructure, low bit rate limitations through gateway implementing LoRa technology at the layer.
- ***Sigfox*** is similar to LoRaWAN, the wireless technology with reduced bitrate and low-cost infrastructure reflects LPWAN technology. It connects devices of large distances by using star topology. This technology functions under unlicensed frequency bands, managed by the organization.
- ***Narrowband IoT (NB-IoT)*** is one of the advancing wireless technology, categorized as LPWAN. It offers many numbers of systems with a licensed spectrum with a low bit rate of specified 3GPP under the single base station.
- ***Power Line Communication (PLC)*** represents a smart grid framework. It is a wired technology that tampered with the impairments related to the wireless media. Its applications are applied to have an innovative smart home living with a low bit rate. IEEE 1901.2 or ITC-TG.9903 are the variants of PLC.
- ***Master-Slave/Token Passing (MS/TP)*** determines the wired technology usually related to the BAC net. This technology is mostly designed based on grid-powered to develop automated features in the physical layer. These are constrained to RS-485 requirements at low bit rates.

**Protocol Architecture for Constrained Node Network (CNN)**

Constrained node network architectures were classified into IP-based architectures and non-IP-based architectures. The main protocol and connections were generated and based on the physical layer.

IP-based architectures are interoperable and flexible to communicate remotely with many devices, sensors, internet connectivities, and actuators. The data that has been shared between devices are secure by providing address names. Any connections to the constrained devices use running IP to obtained efficient internet connectivity. It is an open protocol modeled to have underlying technologies and stack protocols

like HTTP, TCP, IP as frequent network interfaces. This can be mostly observed with grid-powered devices. Some devices with insufficient computational power cannot afford to use traditional protocol stack as it is energy-constrained. To advance more in technology IP based architectures added Ipv4 and then later Ipv6 for system self-configuration. IoT stack-based protocol composes of adaption layer down to IPv6, Low powered routing protocol, Constrained Application protocol, and RPL. This also supports IEEE 802.15.4 networks.

**Network Topology of CNN**
Cities that are developed as smart environment cities adopt mesh network topology. Network with mesh topology is a sophisticated network similar to star topology connections. CNN implements dynamic routing protocol which enables the paths of the network through routers for real-time logical networks. The foremost prominent preferences of the mesh topology of CNN has two advantages. It overcomes the limitations of star topology by reducing the link range. It provides path diversity and avoids star topology single-point failure issues. Due to multiple network connections and path propagation, this will improve the wireless network functionality. This can be applied to develop both smart homes and smart factories with advanced sensors and also with low infrastructure costs. In some cases, LPWAN technologies were also implemented for innovative smart cities.

## *3.3 Smart Environment Sensors with the Internet of Things (IoT)*

The following are the different categories of sensors, technologies that are used to invoke a smart environment using IoT-based Systems. Specialized sensors for smart environments adapt to physical parameters. These sensors are placed for the automatic generation of the information ordered at locations such as homes, body, factory, city, etc. Despite the exploitation of the user information through a smartphone, sensors provide information even though the sensors have not been intended for an unambiguous use case that implicates many disputes.

Specified systems for developing authentic, robustness, and analyzed management of smart factory always encourage to have more sensing effectiveness of the environment, mostly smartphones.

The development of smart homes influences the existence of smartphones. Smart manufacturing devices offer both analog and digital sensors to have quick look over work status. Some of the most implemented manufacturing applications are Industry 4.0 automation. Anticipates the operations that supply information from the device or production of the analyzing process generates a high quality of sensors. Which allows us to have smart homes. Some of the privacy preservations for sensors are considered to come against environmental disputes and performance such as a change in the moisture, the existence of chemicals, vibrations, changes in temperature, dust.

Smart environments mostly reflect the design of smart homes and also overcome the changes affected by nature, i.e., high temperatures, and shocks using a large number of available technologies. Smart homes use ITUTG.9959, infrastructure uses PLC and DECT-ULE. Commonly used technologies are IEEEc802.15.4, BLE. Smart factories use IEEE 802.15.4e TSCH technologies leads to having primary settings. Emerging technologies of LPWAN uses LoRaWaN, NB-IoT, Sigfox. Eventually, the physical environment carries e-health smart applications as home-centric connected to smartphones.

## 4 Conclusion

Remote operations using IoT-enabled keen cloud computing frameworks provide smart homes, advanced healthcare, smart manufacturing factories, and cities. However, selecting the correct innovative technology reaches the best prerequisites of a particular framework with a challenging errand and produces the best framework modeler due to the huge differing qualities of preferences. The trade of information permitted by IoT innovations and frameworks could be the beginning state for enhancing smart environments encouraged by manufactured and encompassing insights. This paper concludes to design the core requirements that generate reusable features and technologies to develop a smart environment. Two technological architectures like Radio Frequency Identification (RFID) and Constrained Node Network (CNN) are described to enhance the Internet of things. Thus the process of inventing a smart environment for the social changes the style of living to advanced good health and safe life.

## References

1. Silva BN, Khan M, Han K (2018) Towards sustainable smart cities: a review of trends, architectures, components, and open challenges in smart cities. Sustain Cities Soc 38:697–713. https://doi.org/10.1016/j.scs.2018.01.053
2. Bellavista P, Chessa S, Foschini L, Gioia L, Girolami M (2018) Human-enabled edge computing: exploiting the crowd as a dynamic extension of mobile edge computing. IEEE Commun Mag 56(1):145–155. https://doi.org/10.1109/MCOM.2017.1700385
3. Castro-Jul F, Diaz-Redondo RP, Fernandez-Vilas A (2018) Collaboratively assessing urban alerts in ad hoc participatory sensing. Comput Netw 131:129–143. https://doi.org/10.1016/j.comnet.2017.12.008
4. Cesta A, Cortellessa G, Fracasso F, Orlandini A, Turno M (2018) User needs and preferences on AAL systems that support older adults and their carers. J Ambient Intell Smart Environ 10(1):49–70. https://doi.org/10.3233/ais-170471
5. Chin J, Callaghan V, Ben Allouch S (2019) The Internet of Things: reflections on the past, present and future from a user centered and smart environments perspective. J Ambient Intell Smart Environ 11(1)
6. Cook DJ, Duncan G, Sprint G, Fritz RL (2018) Using smart city technology to make healthcare smarter. Proc IEEE 106(4):708–722. https://doi.org/10.1109/JPROC.2017.2787688

7. Dubois, Charpillet F (2017) Measuring frailty and detecting falls for elderly home care using depth camera. J Ambient Intell Smart Environ 9(4):469–481. https://doi.org/10.3233/ais-170444
8. Gams M, Gams IM, Yu-Hua Gu I, Harma A, Menoz A, Tam V (2019) Artificial intelligence and ambient intelligence. J Ambient Intell Smart Environ 11(1)
9. Gomez C, Paradells J, Bormann C, Crowcroft J (2017) From 6LoWPAN to 6Lo: expanding the universe of Ipv6-supported technologies for the Internet of Things. IEEE Commun Mag 55(12):148–155. https://doi.org/10.1109/MCOM.2017.1600534
10. Jean-Baptiste E, Russell M, Howe J, Rotshtein P (2017) Intelligent prompting system to assist stroke survivors. J Ambient Intell Smart Environ 9(6):707–723. https://doi.org/10.3233/AIS-170461
11. Lv Z, Li X, Wang W, Zhang B, Hu J, Feng S (2018) Government affairs service platform for smart city. Future Gener Comput Syst 81:443–451. https://doi.org/10.1016/j.future.2017.08.047
12. Masera M, Bompard EF, Profumo F, Hadjsaid N (2018) Smart (electricity) grids for smart cities: assessing roles and societal impacts. Proc IEEE 106(4):613–625. https://doi.org/10.1109/JPROC.2018.2812212
13. Preuveneers D, Ilie-Zudor E (2017) The intelligent industry of the future: a survey on emerging trends, research challenges and opportunities in Industry 4.0. J Ambient Intell Smart Environ 9(3):287–298. https://doi.org/10.3233/AIS-170432
14. Streitz N (2018) Beyond 'smart-only' cities: redefining the 'smart-everything' paradigm. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-018-0824-1
15. Streitz N, Charitos D, Kaptein M, Bohlen M (2019) Grand challenges for ambient intelligence and implications for design contexts and smart societies. J Ambient Intell Smart Environ 11(1)
16. Tripathy K, Tripathy PK, Ray NK, Mohanty SP (2018) iTour: the future of smart tourism: an IoT framework for the independent mobility of tourists in smart cities. IEEE Consumer Electron Mag 7(3):32–37. https://doi.org/10.1109/MCE.2018.2797758

# A Review of Healthcare Applications on Internet of Things


**S. Chitra and V. Jayalakshmi**

**Abstract** Internet of Things (IoT) is rapidly advancing as a most recent exploration point in various academic and industrial controls, especially in medical care, as the flow transformation of the Internet. Due to the rapid proliferation of wearable devices and smartphones, innovation facilitated by the Internet of Things is evolving medical treatment from the conventional center-based setting to a more Personalized Healthcare System (PHS). Nevertheless, making the use of cutting-edge IoT innovation in PHS remains essentially testing within the area thinking about various problems, such as lack of functional and accurate clinical sensors, unstandardized IoT system models, heterogeneity of related wearable gadgets, multi-dimensionality of generated knowledge, and interoperability popularity. This paper offers a description of the IoT of various diseases in healthcare systems.

**Keywords** Internet of things · Personalized healthcare · Big data · Sensors · Algorithms

## 1 Introduction

Internet of Things (IoT) has evolved as another data innovation worldview aimed at creating a specific global organizational system by integrating a variety of physical and virtual 'things' with flexible and sensor growth. At first, IoT was suggested to refer to extremely recognizable articles (things) and their interactive portrayals in a web-like framework, with the use of creativity in radio-frequency identification (RFID). Later on, with a range of sensors, such as actuators, global situating framework gadgets, and smartphones, the concept of IoT was reached out to cover more kinds of 'things'. At an Internet-related point, the consistent mix and efficient outfit

S. Chitra (✉) · V. Jayalakshmi
Department of Computer Applications, VISTAS, Chennai, India

V. Jayalakshmi
e-mail: jayasekar.scs@velsuniv.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_19

of these sensors has posed a large number of exploration problems, from framework design, knowledge preparation to applications [1].

The inspiration for the use of current Information and Communication Technologies (ICT) in the medical services context is typically to provide promising responses to the proficient transmission to patients of a wide range of clinical medical care administrations, called e-wellbeing, such as electronic record frameworks, telemedicine frameworks, personalized determination gadgets, etc.

An ordinary case of medical care IoT arrangements is the estimation of vital signs through far-off sensors and the implementation of nonstop and continuous information through cloud administration to the considered expert. Through IoT execution, medical attendants and specialists don't get the chance to stroll around the wards to follow the states of patients. Rather, a point of interest or irregular sign distinguished from the patients may be followed right away. Other than that, the gadgets with wellbeing observing and action following capacities are encountering expanded prevalence, as they permit clients to turn out to be more mindful of their wellbeing related conduct [2].

IoT offers suitable technologies for numerous applications covering all facets of life, such as smart cities [3], smart traffic management, waste management, control of systemic health, protection, emergency response, supply chain, retail, industrial management [4–7] and healthcare. By 2030, 500 billion devices will be connected, which is roughly equal to 58 smart devices per person on our planet, according to a study by CISCO. IoT industry analysis conducted by Statista at the end of 2017 reported that the global market value of IoT will hit USD 8.9 trillion by 2020, and 7% of the overall market value comes from the healthcare sector. With the introduction of IoT and cloud technologies into the healthcare industry, healthcare providers are able to deliver healthcare facilities that are quicker, more effective, and cheaper, resulting in better customer engagement. As a result, better hospital facilities, better customer engagement, and fewer paperwork for health workers are brought in.

Big Data in medical services alludes to such vast and complicated electronic well-being information collections that they are problematic (or difficult) to deal with standard programming or possibly equipment; nor could the board instruments and techniques be efficiently made up of traditional or usual information. Because of its amount and the breadth of information forms, and therefore the pace at which it must be monitored, large-scale information in medical care is daunting. It includes emotionally supportive clinical information and clinical choice networks (doctor's composed notes and medications, clinical imaging, research centre, drug store, protection, and other management information); quiet information in electronic patient records (EPRs); machine-generated/sensor information, as from the verification of critical signs; web-based media posts, including Twitter channel posts And not a lot of patient-explicit data, including information on crisis treatment, news outlets, and clinical journal articles. There is room for the enormous knowledge researcher, among this enormous amount and exhibition of knowledge. Huge data investigation will theoretically improve consideration, save lives, and lower costs by identifying affiliations and having examples and trends within the data [8].

Big Data in medical care requires the processing of large quantities of knowledge from various medical service institutions, followed by the processing, supervision, dissection, analysis, and dissemination of viable interactive data. Six attributes, viz., amount, assortment, speed, veracity, fluctuation, and value, describe the enormous knowledge in medical care. There are numerous large-data logical devices such as Apache (Pig, Spark, Flume, Tez, Oozie, Hive, etc.), Hadoop (HDFS, MapReduce), MangoDB, and so on. Each instrument has its own utility and strength [9].

The remainder of the analysis paper is sorted accordingly. The IoT feature in healthcare structures is audited in Sect. 2. Section 3 addresses the analysis of IoT in different structures of medical care. Section 4 explains the analysis of the examination with its consistency and shortcomings provided in a table. In Sect. 5, the conclusion is given.

## 2 Role of IoT in HealthCare

The aim of the IoT is to disrupt the medical care industry with innovative advances. The worldwide pandemic caused due to COVID-19 has offered degree to new developments in the web of things. IoT-associated medical services applications offer continuous observing and keen clinical IoT gadgets synchronized to a cell phone application that empowers specialists to gather clinical information of their patients at some random spot or time.

In the event of a health-related crisis such as cardiovascular breakdown, diabetes, asthma attacks, and so on, continuous monitoring using associated gadgets will save a million lives. The associated IoT gadget collects and moves information about well-being, such as pulse, oxygen, and glucose levels, weight, and ECG data. The knowledge is put away in the cloud and, according to the sharing access authority, can be imparted to an authorized person or a doctor.

Interoperability, machine-to-machine communications, data trade, and knowledge creation are facilitated by IoT in medical care, rendering the transmission of medical care administration extremely financially savvy. By minimizing superfluous visits and using higher-quality properties, this innovation-driven structure will minimize costs.

In dangerous situations, continuous follow-up and alarms will transform as a hero to protect the well-being of a simple patient with clear reminders and ongoing cautions for valid testing, review, and study.

For a patient looking for professional assistance, it is a terrible situation, but being unable to communicate with a specialist because of impediments, such as region and lack of knowledge. Via medical services conveyance bonds associated with patients through IoT gadgets, patients may take clinical remedies comfortably. IoT medical care applications help doctors practice, forestall, and analyze medication all the more without any problem. With continuous information and the likelihood to break down past medicines and determination of a patient—Smart medical care frameworks utilizing IoT assists with decreasing mistakes. Besides, consistent robotized

observing and upgraded investigation of the patient's condition prompt appropriate treatment without the chance of mistake.

The information accumulated by IoT medical services gadgets are exceptionally precise and empowers the specialists to settle on educated choices. Understanding history can be examined and estimated quickly. Information can likewise be sent to a leading body of specialists or medical care experts on a cloud stage.

## 3 Literature Review

Versatile advancements these days assume basic parts in medical care checking and benefits. These advances incorporate cell phones, individual computerized associates (PDAs), portable cameras (e.g., SenseCam), savvy watches, and so forth. As the vast majority of cell phones are installed an assortment of inertial sensors (e.g., accelerometer, spinners, and so forth) and biomedical sensors (skin temperature, pulse, and so on), they are intended for giving customized and nonstop cares to clients. This segment intricately talks about the writing survey on how IoT has been dealt with various constant sicknesses.

Earlier, Al-Makhadmeh and Tolba [10] introduced an IoT-focused clinical gadget for social events, followed by HD patient heart status results. The data was prepared in the wake of being relentlessly communicated to the HC emphasis using the higher-request Boltzmann deep neural organization conviction. The deep learning process, which prompted the strong control of unpredictable data to achieve efficiency, discovered the previous HD angles. The framework's exhibition was processed with an emphasis on these attributes [particularity, f-measure, misfortune function, affectability, alongside collectors working qualities (ROC)]. This current method had 99.03% accuracy with a minimum time complexity of 8.5 s successfully; minimal HD mortality was achieved along these lines by reducing the multifaceted architecture in HD diagnosis.

Majumder et al. [11] implemented a different tactile system by using a keen IoT that gathered information from the Body Area Sensor to provide early provision for fast approaching heart failure. The aim behind the current work was to construct a synchronised brilliant IoT that integrated a correspondence unit of lower power so that one could unnoticeably accumulate pulses using a cell phone alongside internal heat levels without blocking their normal daily life. For sensor information review for detecting and predicting abrupt coronary failure with greater precision, the sign preparing alongside ML techniques was presented.

Mohan et al. [12] suggested a method that discovered big highlights by implementing ML techniques to boost the precision of the cardiovascular disease conjecture. By joining the singularity of the RF couple with the Linear Method (LM), the half and half RF with a direct model was used. This is set up in the HD figure to be very accurate. By means of the forecast model planned for HD, this approach gave an enhanced exhibition level with an 88.7% accuracy level.

Miramontes et al. [13] have planned the PlaIMoS fixed estimation station which estimates persistent blood oxygen immersion, breath rate, and galvanic opposition, which are distinguished by Radio recurrence recognizable proof (RFID) sensor and a PlaIMoS far off hub that gets understanding blood temperature, electrocardiogram, beat rate, and fall information. The fixed estimation station is associated with the web by methods for its Wi-Fi interface which sends all data onto the PlaIMoS information base, situated inside the cloud. The PlaIMoS distant hub utilizes the IEEE 802.15.4 norm to communicate information to the WSN foundation. The Web administration gathers the data from the WSN framework and communicates the information to the PlaIMoS data set. All data put away inside the information base is used by the PlaIMoS API (Web administration) and introduced in JavaScript object documentation (JSON) design for the different applications. The iOS, Android and Windows applications show the information to the taking an interest patients, specialists, and overseers and push cautions when the clinical boundaries surpass the extents set up by the specialist.

Tan and Halim [14] present the AI system to consequently foresee the potential dangers, for example, diabetes and kidney malady. The system measures three fundamental main signs, such as internal heat intensity, beat rate, and heartbeat, planning and adjusting sensor signs to meaningful yield with high precision and displaying the observation results on web apps, for example. The medical services system infrastructure structure is split into a front-end (introduction layer) and a back-end (information access layer). The system coordinates microcontroller-powered sensors throughout the front-end layer to measure a client's critical signals and uses Arduino Nano and Intel Edison as discernible yields to move across the known indicators. The front-end layer contains an interactive UI (GUI) or software application that helps the two patients and specialists to navigate and display the information and data collected. The yield from the information procurement system is imposed on the cloud data collection in the back-end layer and the prior knowledge of information is analyzed. The synchronized sensors are attached to the Intel Edison point, and the yield readings are sent to IBM Bluemix for the database and display of cloud information. The consistency of the model produced is 90.54% and 87.88% respectively, between a stable human and patients with diabetes and kidney failure.

A patient behavior observing system of subjective examination for cardiovascular disease ID and dropping detection was introduced by Chui et al. [15]. It is an enormous knowledge and IoT-based patient behavior control system that can be extended to various applications I malignant growth medicines and genomics; (ii) alert and alarm for quiet detection of vital signs; (iii) medication viability review; (iv) market security of medical care; and (v) prevention and location of extortion. The patient data floods that will be absorbed into Kafka, which is a celebrated open-source, low dormancy, high performance, and scalable programming stage for stream planning. Breathing rate, pulse, circulatory tension, and internal heat level sensors are above all controlled by the mill's clinical and body sensors, as are electrocardiogram, electroencephalography, electrooculography, electromyography, accelerometer, cardioverter defibrillator, and pacemaker. The number of sensors associated with each patient depends on the use, normally directed by specialists in medical services. Segment

statistics, safety data, study centre findings, and clinical records are contained in the very own data of the patients. Third, data logging allows all agreements to be pursued by details, documents or applications that are modified, obtained, or placed away on capability applications or gadgets.

A portable gadget with a bio-detecting facial cover is designed by Yang et al. [16] to track a patient's tormenting strength using facial surface electromyogram (sEMG). The wearable gadget functions as a remote sensor hub and is fused for distant agony observation into an Internet of Things framework. Up to eight channels of sEMG could be investigated at 1000 Hz within the sensor centre, to allow the maximum recurrence spectrum, and gradually transmitted to the cloud worker through the entryway. Expansion of that, through the portable gadget strategy for long-term monitoring, both low vitality use and wearing solace are regarded. A compact web application is created for persistent gushing of high-volume sEMG information, advanced sign planning, deciphering, and representation to detail endless pain information to guardians distantly. As an extension between the sensor centre and the software, the cloud stage within the system supervises interactive communication with the worker and the web application. In summary, this research proposes an adaptable IoT system for continuous biopotential regulation and a wearable reaction by external appearances for programmed torment assessment.

Using Arduino Uno and GSM/GPS technology, Veyilazhagan and Bhanumathi [17] have developed a health surveillance device to protect sick patients and elderly people by an embedded system. To successfully maintain the system, sensors such as temperature, gas, drop detection, heart rate, and blood pressure rate are used. As the central diagnostic tool, this control system was primarily developed for cardiac treatment with electrocardiogram (ECG) signal analysis. The machine consists of three key components, namely (a) a mobile gateway, a mobile computer that collects 12-lead ECG signals from every ECG sensor, (b) a remote server component that hosts algorithms for precise annotation and interpretation of the ECG signal, and (c) a doctor's point of care computer to obtain a medical diagnostic report from the server hardware on the server hardware. For modeling and visualization, the Proteus virtual instrument is used. The GSM module sends SMS updates and health information to the doctor's cell phone.

## 4 Comparison Table

The comparison table for the techniques reviewed, sensors, devices, network, and data set used along with strength and weakness of the related work is evinced in Table 1.

**Table 1** Comparison table on techniques, devices, sensors, network, and data set used in survey

| Year | Techniques | Sensors used | Device | Network | Data set | Strength | Weakness | Reference |
|---|---|---|---|---|---|---|---|---|
| 2019 | HOBDBNN Approach, Deep learning | ECG, HR, BP sensors | Wearable watch | Bluetooth | UCI Machine Learning Repository data set | most extreme acknowledgment precision and least time intricacy Limited the HD mortality | The calculation didn't utilize any advancement calculation focused on include choice, which expanded the preparation season of the calculation and that raised a few troubles in dataset the executives for forecast | [10] |
| 2019 | Multisensory utilizing a system with smart IoT | PR, Temp sensor | Arduino Uno$^{TM}$, smart phone, wrist band | Bluetooth | Public data sets | Mix of sign handling alongside ML calculations viably recognized the unexpected cardiovascular failure with higher exactness | The calculation didn't bring up information security notwithstanding the framework isn't savvy | [11] |
| 2019 | HRFLM | HB, Temp, humidity sensor | Arduino/Genuino, smart phone | GSM/Bluetooth | UCI Cleveland Repository | Having the improved presentation level with 88.7% precision level | The framework didn't have the ability of observing the HD continuously | [12] |

(continued)

**Table 1** (continued)

| Year | Techniques | Sensors used | Device | Network | Data set | Strength | Weakness | Reference |
|---|---|---|---|---|---|---|---|---|
| 2017 | Not mentioned | Airflow, Temp, HR, ECG, BO, respiration rate, galvanic sensors | Rasperry Pi | Wi Fi | Public data set | PlaIMoS stage performed well in an emergency clinic setting, safely checking, and detailing understanding data progressively conditions | The framework didn't cross allude referenced to give a top to bottom symptomatic to clinical professionals | [13] |
| 2018 | Machine learning | Temp, BP, PR sensors | Arduino Nano and Intel Edison | Wi Fi | UCI ML repository | The investigation presents system to naturally foresee the potential dangers, for example, diabetes and kidney infection | The framework introduced examination are a fundamental examination that requires further clinical test and constant information to additionally approve the ML model before framework coordination | [14] |

(continued)

**Table 1** (continued)

| Year | Techniques | Sensors used | Device | Network | Data set | Strength | Weakness | Reference |
|---|---|---|---|---|---|---|---|---|
| 2019 | Machine learning | BR, PR, BP, Temp, ECG sensors | Not mentioned | Wi Fi | Not mentioned | The system can be applied for various applications as cancer treatment, genomics, patient vital sign monitoring, efficiency of treatment, fraud prevention and detection | The framework needs true contextual investigation and restricted subjective examination on just two patient social observing applications | [15] |
| 2017 | Big data processing and deep learning | sEMG sensor facial mask | Not mentioned | Wi Fi | Public data set | Low vitality utilization and agreeableness expand the capability of detecting facial cover and sensor hub for long-haul measure | The system needs to test on large data set | [16] |
| 2017 | No algorithm | Temp, Gas, Fall, HR, BP, respiration, PR sensors | Arduino Uno | GSM/GPS | Not mentioned | This system is implemented with the low cost to save life in emergency conditions | The system did not apply any optimization algorithm and it needs to be tested on real-time data | [17] |

*HR* Heart rate, *BP* Blood pressure, *PR* Pulse rate, *Temp* Temperature rate, *HB* Heart beat, *BO* Blood oxygen, *BR* Breathing, *ECG* Electrocardiogram

## 5 Conclusion

The worldview of the Internet of Things speaks to the vision of the subsequent ICT transition sprint. IoT-enabled PHS advancement would allow faster and safer preventive evaluation, lower total costs, increased patient-focused practice, and improved managementability. IoT empowered PHS can possibly fortify our way of life in numerous different viewpoints. In this overview, We checked on the overarching best-in-class innovations for IoT-empowered medical services applications from different points of view. We examined current innovation, sensors utilized, and preparing procedures. All the more critically, we gave an elevated level depiction of differed IoT empowered medical care applications. However, the objectives discovered for IoT in medical services are not effectively reachable, and there are as yet numerous difficulties to be confronted and, thusly, this examination field is getting increasingly more impulse. Therefore, it can be summarized that the entire idea of the IoT and its complete implementation in medical treatment and human prosperity will come from this synergistic approach.

## References

1. Qi J et al (2017) Advanced internet of things for personalised healthcare systems: a survey. Pervasive Mob Comput 41:132–149
2. Shin DH, Lee S, Hwang Y (2017) How do credibility and utility play in the user experience of health informatics services? Comput Hum Behav 67:292–302
3. Dang LM, Hassan SI, Im S, Moon H (2019) Face image manipulation detection based on a convolutional neural network. Expert Syst Appl 129:156–168
4. Nguyen TN, Thai CH, Luu AT, Nguyen-Xuan H, Lee J (2019) NURBS-based postbuckling analysis of functionally graded carbon nanotube-reinforced composite shells. Comput Methods Appl Mech Eng 347:983–1003
5. Nguyen TN, Thai CH, Nguyen-Xuan H, Lee J (2018) NURBS-based analyses of functionally graded carbon nanotube-reinforced composite shells. Compos Struct 203:349–360
6. Nguyen TN, Lee S, Nguyen-Xuan H, Lee J (2019) A novel analysis-prediction approach for geometrically nonlinear problems using group method of data handling. Comput Methods Appl Mech Eng 354:506–526
7. Dang LM, Hassan SI, Im S, Mehmood I, Moon H (2018) Utilizing text recognition for the defects extraction in sewers CCTV inspection videos. Comput Ind 99:96–109
8. Saranya P, Asha P (2019) Survey on big data analytics in health care. In: Proceedings of the 2nd international conference on smart systems and inventive technology (ICSSIT), vol 5, pp 46–51
9. Senthilkumar SA (2018) Big data in healthcare management: a review of literature. Am J Theor Appl Bus 4:57
10. Al-Makhadmeh Z, Tolba A (2019) Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: a classification approach. Meas J Int Meas Confed 147
11. Majumder AJA, Elsaadany YA, Young R, Ucci DR (2019) An energy efficient wearable smart IoT system to predict cardiac arrest. Adv Hum-Comput Interact 2019
12. Mohan S, Thirumalai C, Srivastava G (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 7:81542–81554

13. Miramontes R et al (2017) PlaIMoS: a remote mobile healthcare platform to monitor cardiovascular and respiratory variables. Sensors (Switzerland) 17:1–24
14. Tan ET, Halim ZA (2019) Health care monitoring system and analytics based on internet of things framework. IETE J Res 65:653–660
15. Chui KT, Liu RW, Lytras MD, Zhao M (2019) Big data and IoT solution for patient behaviour monitoring. Behav Inf Technol 38:940–949
16. Yang G et al (2018) IoT-based remote pain monitoring system: from device to cloud platform. IEEE J Biomed Health Inform 22:1711–1719
17. Veyilazhagan R, Bhanumathi V (2017) An outdoor intelligent health care patient monitoring system. In: 2017 international conference on innovations in green energy and healthcare technologies (IGEHT). https://doi.org/10.1109/igeht.2017.8094061

# Big Social Media Analytics: Applications and Challenges


**Sonam Srivastava and Yogendra Narain Singh**

**Abstract** Social media plays an indispensable role in the ever-expanding human population, which will in turn result in generating a huge volume of data. Social media analytics is the technique used for acquiring and analysing the data from social networks. The big data associated with social media finds state-of-the-art applications in the distinctive socio-economic domains. A considerable research literature evidence is available on the challenges that are inherent in particular applications of social media or big data separately, but there is hardly any exclusive study that has been performed on the social media big data. To address this gap, this paper presents the state-of-the-art social media big data applications in business, healthcare, education and crisis management with which the challenges associated with them are also critically evaluated. Also, different frameworks of big social media analytics are presented, and their potential in addressing the application-specific challenges is also described. Henceforth, this research article leverages significant advantages to researchers and experts, who wants to gather and analyse the social media data.

**Keywords** Social media · Big data · Social media analytics · Business · Healthcare · Education · Crisis management

## 1 Introduction

Today, the amount of data is undergoing unprecedented changes as a result of advances in Web technology, social media, smart phones and sensor-based devices. There are over 3.9 billion social media users worldwide, and their penetration rate is about 51% [1]. Big data market report estimates that the global size of the big data industry will rise from 138.9 billion USD in 2020 to 229.4 billion USD by 2025

S. Srivastava (✉) · Y. N. Singh
Institute of Engineering and Technology, Dr. APJ Abdul Kalam Technical University, Lucknow, Uttar Pradesh 226021, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_20

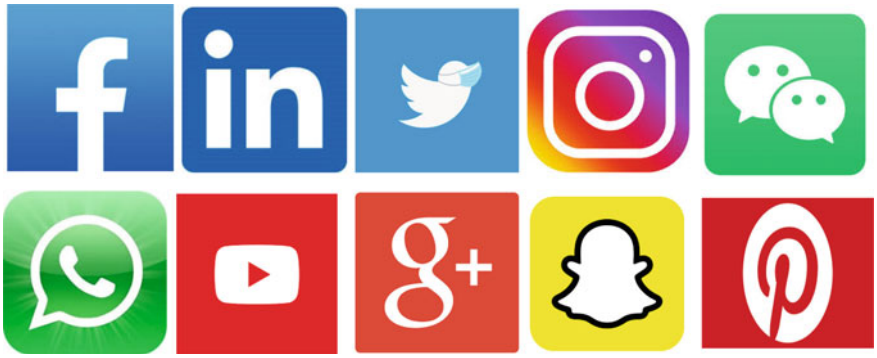


**Fig. 1** Social media platforms

[2]. Big data deals with different ways of analysing, systematic data extraction and dealing with data sets that are too complex to handle the conventional data processing applications.

Social media is generally Internet-based applications that are used for sharing the information about social, economic and political interests among virtual communities [3, 4]. The user contents, *i.e.*, comments, textual post, videos and digital photographs, that are generated through the online interactions are remaining as the impulse of social media. The prominent resources of social media include Facebook, LinkedIn, Twitter, Instagram, WeChat, WhatsApp, YouTube, Google+, Snapchat, Pinterest, *etc.*, as shown in Fig. 1. The social media provides a range of benefits and opportunities to empower both people and communities in a variety of ways. Social media has also become a crucial handler for the acquisition and dissemination of information across various domains such as science, entertainment, business, politics, academics, healthcare and crisis management [5]. The social media has become an integral part of digital society with the proliferation of social websites and applications, thus resulting in explosion of data during the recent years. The social media big data holds the characteristics of big data, *i.e.*, it is voluminous, structured or unstructured and generated in real time or streamed. The data generated from social media can be imprecise and uncertain, while issues of information security and integrity are the prime ones [6].

The enormous growth of social media usage has led to an increasing accumulation of data called social media big data. This data may be of unstructured or structured nature [7]. The textual content of social network is an example of unstructured data, while the friends' relationship is the examples of structured data. The social media big data is highly dynamic and complex in nature. It may be incomplete and inconsistent due to availability of unreliable internet sources. Unlike traditional data processing and data mining techniques, the available social media big data needs processing and

knowledge discovery. These social media big data possess enormous market value potential in distinct fields of transport, genetics, online advertisements, finance and energy management. Traditional methods, however, are failing in the face of such large data.

In the recent past, a significant interest is shown in analysing voluminous social media data ranging from business interests to societal benefits. This paper presents a critical survey of big data social media applications in business, healthcare, education and crisis management. This research work critically evaluates each domain of application and the associated challenges with the growth of digital data from its current volume of 44 zettabytes (ZB) to 175 ZB by 2025 [8]. This work also presents the application-specific frameworks of social media analytics to address their challenges, for example, the identification of potential customers, opportunities for targeting advertising, social customer relationship and business intelligence for marketing in business applications [9, 10]. The sharing of medical information with practitioners, feedbacks, health alerts and awareness programmes in pandemic like COVID-19 affects the patient care quality in healthcare applications. The sharing and exchange of academic information to their stake holders make the education convenient and exploration of improved learning methods in education application. The social media data also helps in minimising the losses due natural calamities while defending state or organisational reputation in crisis management applications.

The rest of the paper is organised as follows. The literature review of big social media data analytics is given in Sect. 2. The applications of social media big data along with their challenges are presented in Sect. 3. The description of frameworks of social media analytics to different applications are given in Sect. 4. Finally, the conclusion is drawn in Sect. 5.

## 2 Literature Review

The big data transformation aims to change the conventional way of how humans work, function and think by allowing process optimisation, enabling the exploration of knowledge and enhancing decision-making. In the context of big data, Heureux et al., have presented a comprehensive analysis of the challenges associated with machine learning and classified them according to the four major Vs of big data, *i.e.,* volume, velocity, variety and veracity [11]. Sivarajah et al., have conducted a comprehensive review that presents a futuristic view of big data challenges and big data analytics. Their study assists researchers to build new solutions on the basis of the challenges that they spotted. Stieglitz et al., have presented the issues of social media analytics, conducted literature analysis and also suggested solutions for the addressed challenges [12]. Their work focuses on the major steps involved in social media analytics, *i.e.,* data discovery, collection, preparation and analysis. The findings may benefit to the researchers working in the area of social media analytics by providing them an insight to solve the issues for social welfare.

Oztamur et al., have intended to examine the role of social media with respect to small- and medium-sized enterprises. They considered social media as a novel strategic tool for social network marketing [13]. Their research focuses on analysis and comparison of American and Turkish companies' accounts on social media. In order to analyse social media data for understanding students learning patterns, a work flow is presented integrating qualitative analysis and mining techniques [14].

Nowadays, social media platforms can be used to track diseases. Alessa et al., have claimed that social media and search engines can be used to predict the seasonal epidemics faster than the government agencies [15]. They have reviewed prevailing alternate solutions that track influenza outbreak using Web blogs and social networking sites. A similar research has been conducted by Qin et al., in the context of pandemic using social media search index [16]. It used to predict the number of novel coronavirus, *i.e.,* COVID-19 cases. The authors have claimed that social media search index is an effective predictor to find the correlation with newly confirmed COVID-19 cases. The methodology for crawling, processing and filtering the tweets accessible freely by the public has been presented in [17]. They used REST API to acquire tweets from Twitter in real time. The authors asserted that mining tweets have potential to supplement current traffic incident data in an inexpensive manner. The concept of crisis and key developments that affect the sense of crisis management around the world is presented in [18]. The conceptualisation of crises, their management and the effects on growing use of social media and connected devices is studied.

## 3 Applications of Social Media Big Data

It has been reported that in the currently existing digital universe, where 49% of worldwide data would be based on the public cloud environment by 2025 [19]. Nearly 30% data of the world will require real-time processing by 2025. By now, there are over 5000 GB data shared to each person that needs to be processed and discover knowledge from them. There are less than 1% of worldwide data which is analysed, and less than 20% of this data is protected. The sources of this data are business and advertising, healthcare, academics, politics, environmental and social media platforms including machine-generated data. The different applications of social media big data are described, for *e.g.,* business, healthcare, education and crisis management along with their challenges in this section.

### *3.1 Business*

Social media finds its importance in marketing products, promoting brands, connecting to new customers and fostering new business. It provides information on the hobbies, demographics, habits and characteristics of social media users. Social media

marketing takes the benefit of social networking to assist the firms to enhance brand exposure and widen customer reach [20]. It utilises the features of social media, *i.e.*, online communities, and social data in order to take effective business decisions [21]. Social media data enables the advertisers to determine how current and potential customers perceive their products in real time [22, 23]. Thus, it is evident that the social media helps the organisations to acquire customers feedback related to their product which can be used to improve decisions and get value out of their business. It requires careful thinking and planning to use social media for getting the optimum benefit out of business.

In order to resolve the challenges, it is necessary to recognise and leverage the business services provided by social media platforms. Social media makes it simpler for businesses to target consumers directly, create brand awareness, promote products and interact directly with current and future clients. Although, social media enables consumer loyalty, engaging in a public discussion platform often poses risks. There must be a clarity on how negative reviews, inappropriate language and libellous material about the organisation should be handled. It is challenging to ensure that whatever content is shared on social media and how one communicates with people conveys a professional image to the world. Preparing a set of guidelines to handle social media would help to steer through the challenges.

To address these issues, it is necessary to understand the legal consequences and best practise using social media. For example, before launching any website, the insurance about legal aspects like privacy policy, disclaimers, terms and conditions need to be understood.

## *3.2 Healthcare*

In 2018, 2.65 billion individuals worldwide used social media. This number is expected to rise to 3.1 billion by 2021 [24]. Contemporary lives are going digital, and healthcare is not an exception. Social networks have emerged as an essential health resource not only for millennials but also older generations those utilised social media for seeking and sharing health information. Over 99% hospitals across USA have their own active Facebook page, and a rising number are also establishing their existence on other social media platforms like Instagram and Twitter. Through these platforms, the health systems and physicians receive feedbacks, share health alerts and strengthen their brand with focus on creating faith. Medical health researchers have utilised Twitter to identify the publics' behaviour towards disputed health subjects such as E-cigarettes [25]. Medical field mines social media data to rapidly determine the outbreaks of contagious diseases and to circulate useful information in suppressing the spread of these diseases [26, 27].

In the digital era, empowered and digitally-enabled customers want more value than just 'social' from social media. They are searching for meaningful information that educate, inform and help them for finding resources and make more objective health care decisions. The interests of social media healthcare data are protected by

Healthcare Insurance Portability and Accountability Act (HIPAA) [28]. The common violations of this act include the posting of data, diagnostic reports, photograph and videos that could enable the identity of patients without their consent.

The demand of patients with the trust of healthcare system to keep up with social trends having a social media presence for smart care. For example, 65% of Americans attempt to self-diagnose health problems using the Internet instead of simply consulting health professionals. The results of their searches induced stress in 74% of these individuals [29].

### *3.3 Education*

Social media enables the students to get more useful information and connect with educational systems and other learning groups to make education favourable. Social media has gained eminence as a credible source of information where educational institutions can link with their stakeholders globally. Studies have revealed that 96% of students have access to Internet and at least one social media site [30]. Students exchange valuable information about their examination and classes through personal computer, laptops, tablets and mobile phones. They also transmit tips, study materials, views, school projects and several kinds of valuable reading stuffs each another. Many universities and colleges have completely embraced social media networks, *e.g.,* Facebook, YouTube and Twitter, which are connected with large number of students [31]. These channels are convenient in publishing school news and educational information, make declarations and tackle various issues related to students through group associations. It also serves as a tool for the benefit of the students where they establish connections among stakeholders of academia.

However, social media also has an adverse impact as it affects relationship with others, people get distracted easily, and might cause cyber bullying [32]. There is a concern that students will be caught up in the chaos of social media rather than concentrating on their work [33]. It could be detrimental to their mental health to place young students on social media platforms for the sake of education. Based on the survey of teens and young adults, an increased use of social media can cause depression, anxiety and loss of self-esteem. Such problems can impact how an individual functions normally and can be harmful to education. If a student is overcome by heavy social media use and mental health problems, then it will be difficult for them to concentrate on their studies [34].

The use of social media in higher education has its drawbacks, such as educator dominance in student–teacher interactions, privacy issues and discriminatory behaviour [35]. The enormous amount of heterogeneous data are created by several universities and other educational sectors. Educators, students, instructors, tutors, researchers, developers and people who deal with educational data are also challenged by the velocity of different data types. Institutes that handle streaming content such as click streams from websites need to update information in real time to serve

the relevant advertising and the best deals to their consumers. Thus, social media has a dual influence on student achievements, and it is imperative to catch up with adolescents use of social media network with prime responsibility.

### *3.4 Crisis Management*

Social media is the two-edged sword in crisis management. Platforms such as Facebook, Instagram and Twitter can be the critical networking tool for handling crisis efficiently. But, if they are not used carefully, then they can also create a crisis worse than ever. It is often utilised as a means of emergency management by informing people about the current status of crisis in the affected area or what actions are to be taken [36]. Various natural disasters such as forest fires, floods, earthquakes and major outbreaks of diseases causing severe symptoms such as pandemic influenza, ebola and corona viruses are experienced at distinct places and around the globe [18]. These dramatic events are communicated, documented and understood via social media using smart devices by citizens, leaders, observers and official responders.

The accumulated social media data can be analysed for spotting a specific location where the traffic accidents occur in real time. The location can be derived from the data by applying named entity recognition method, geocoded tweets or global positioning system [17]. Law enforcement has begun to utilise social media content to solve crimes. To prevent crime, the ATHENA system was developed to assist law enforcement [37]. Accenture collaborated with Singapore government and developed solution for predicting crowd behaviour and potential responses to incidents during 50th anniversary celebration of independence which gave 85% accurate results. Another example includes Kumbh mela experiment that predicted crowd behaviour and possibility of stampede [38]. Thus, social media data assists crisis management to provide public safety by predicting potential activities that disrupt public harmony.

However, the speed of communication on social media is a two-edged sword. This makes it possible for negative feedback to spread rapidly in a way that gets impossible for the organisation to control. Therefore, a new issue can be quickly connected to past crises by the media, and thus, awareness of older facts can exacerbate a situation. Social media involvement makes misinformation propagation even more probable. The arrival of social media, however, also implies that an entity has less power over what is said about them, which can impact their reputation.

## 4 Big Social Media Analytics

Big social media analytics is the synthesis of online behaviours of the users of social websites such as Facebook, WhatsApp, Twitter and LinkedIn. The proliferation of social media applications provide immense opportunities and challenges for researchers. The large data generated by users of social media is the result of their

behaviour, feedbacks and routine activities. Therefore, the aim of big social media analytics is to device intelligence strategies that addresses their application-specific challenges. Different frameworks of social media analytics and their potential in addressing the challenges are also described.

## *4.1 Social Media Business Analytics*

The key concept of business is return on investment. Jeremiah Owyang has built a hierarchy of metrics for social media investments [39]. His idea is to combine different metrics for social media, representing different stakeholders in different positions.

The social media business model aims to perform business and predictive analytics tasks for addressing the issues of an organisation as shown in Fig. 2a. The model recognises crucial success factors and key performance indicators for general online communications. The business analytics is performed by managers and employees engaged in social media with the focus on how social media impacts the business. Business metrics are evolved summarising the social media analytics responsible for predicting growth of the organisation. The predictive analytics measures the strategies and outcomes that lead to learn from failures and direct the company into the foreseen future to the best.

Applying the model can result in creating an overall dashboard for the organisation to track its progress. It can be used to identify the indicators and their interaction to assess the performance in corporate world.

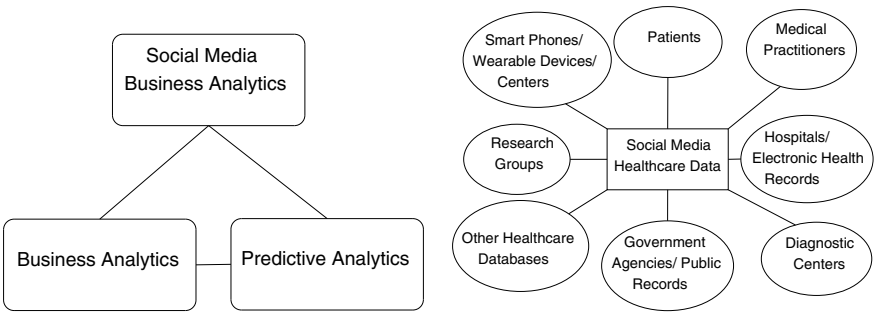



**Fig. 2** Big social media **a** business analytics [40] and **b** healthcare analytics

### *4.2 Social Media Healthcare Analytics*

The social media healthcare analytics objectives are to increase the accessibility, communication and interaction between health practitioners and patients using social media subscriber-based platform for improving healthcare delivery to society. In a normal scenario as shown in Fig. 2b, the patient is registered with the system using social networks. The patients can communicate within their peer group members those are formed on the basis of same disease and symptoms. The findings of these communications are prepared using data analytics techniques.

The patients are monitored using the data shared by biomedical sensors, and the data is forwarded to the system for arranging medical practitioners for consultation. Under the non-availability of the medical practitioner, the system locates other alternatives such as hospitals available nearby using global positioning system. However, the system continuously monitors the patients whether they are admitted in the hospitals or away from the hospitals.

The presented model illustrates how different factors, *e.g.,* medical practitioners, electronic health records, diagnostic centres, public records, research groups, wearable devices and other healthcare databases, can communicate using social media platforms for better healthcare. Although, the model needs to address the issues of data privacy and protection as healthcare applications are combined with social networks.

### *4.3 Social Media Education Analytics*

Presence is considered to be a central concept of online learning. Using the community of inquiry framework, the presence of students available online can be monitored [41]. The framework suggests that when interdependent dimensions of presence, *e.g.,* social, teaching and cognitive interacts, then students experience a meaning learning. The social media education analytics aims to bridge the gap that exists in online education and to overcome the disconnected students feel when studying online. The schematic of SMEDA is shown in Fig. 3. The model explores the role of social media big data in education using big data analytic techniques. Their interrelationships, *e.g.,* education data analytics (EDA), social media educational data (SMED) and social media data analytics (SMDA), make possible how educational communities communicate, socialise and learn themselves. Finally, the educators explore the opportunities to help students and themselves for life-long learning using social media educational data analytics (SMEDA). Further, the pros and cons of social media in education need to be explored and adopt ways to make sure that a healthy educational environment can be built.

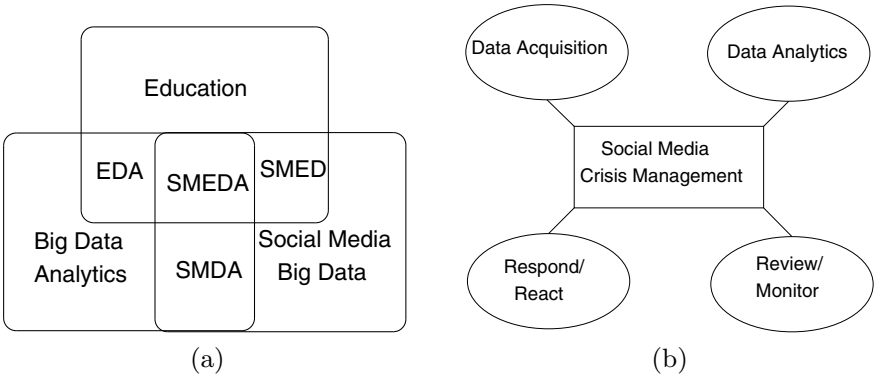



**Fig. 3** Big social media **a** educational analytics [42] and **b** crisis management analytics

### *4.4 Social Media Crisis Management Analytics*

The role of social media analytics in crisis management is very critical in the existing digital universe, where people using social networks are more educated and spoken than their non-users. The intelligent practices of crisis management include to forecast the crises well before the organisation may encounter and suggest appropriate measures.

The social media crisis management analytics model is shown in Fig. 3b. It acquires the information from users of different social media platforms and performs data analytics work to discover the knowledge from them. The model then observes the responses of users and their reactions on an incident belong to social, economical, cultural, political or a natural disaster. Finally, it reviews the reactions of people and suggests the remedial actions to avoid any confusion, restrict the misleading information going viral, and prepare a crisis management strategy to prevent the occurrence of that incident in future.

## 5 Conclusion

Social media is growing rapidly that allows individuals to connect, engage and share ideas online. This paper has presented the state-of-the-art applications of social, economic and political arena. It finds distinctive applications in business, healthcare, education and crisis management. Critical evaluation along with the challenges associated with each application due to exponential growth of social media data has also been done. The different frameworks of social media analytics are presented, and their potential is described for addressing the application-specific challenges.

Social media big data helps companies make more informative business decisions by analysing large volumes of data. Traditionally, by the limited availability of stan-

dardised and consolidated data, the health care industry has faced many setbacks. However, the use of social media in healthcare may extend efficient services to the patients. The use of social networking in education needs an evaluation on student accomplishments. Strategies of crisis management are themselves very challenging, but social media analytics can help in preventing the antisocial activities for maintain public harmony.

## References


1. Social media—Statistics & Facts. https://www.statista.com/topics/1164/social-networks/. Last accessed 10 2020
2. Big Data Market, https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html. Last accessed 11 2020
3. Obar JO, Wildman SS (2017) Social media definition and the governance challenge–an introduction to the special issue. Telecommun Policy 39(9):745–750
4. Kaplan AM (2010) Users of the world, unite! The challenges and opportunities of the social media. Kelley School Bus 53(1):9–68
5. Stiglitz S, Xuan LD, Bruns A, Neubeger C (2014) Social media analytics: an interdisciplinary approach and its implications for information systems. Bus Inf Syst Eng 6:89–96
6. Seo EJ, Park JW (2018) A study on the effects of social media marketing activities on brand equity and customer response in the airline industry. J Air Transp Manage 66:36–41
7. Zeng D, Chen H, Lusch R, Li SH (2010) Social media analytics and intelligence. IEEE Intell Syst 25(6):13–16
8. IDC: Expect 175 Zettabytes of data worldwide by 2025, https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html. Last accessed 14 2020
9. Anshari M, Almunawar MN, Lim SA, Al-Mudimigh A (2019) Customer relationship management and big data enabled: personalization & customization of services. Appl Comput Inf 15(2):94–101
10. Sivarajah U, Irani Z, Gupta S, Mahroof K (2020) Role of big data and social media analytics for business to business sustainability: a participatory web context. Ind Mark Manage 86:163–179
11. Heureux HL, Grolinger K, Elyamany HF, Capretz MAM, (2017) Machine learning with big data: challenges and approaches. IEEE Access 5:7776–7797
12. Stiglitz S, Mirbabaie M, Ross B, Neubeger C (2018) Social media analytics—challenges in topic discovery, data collection, and data preparation. Int J Inf Manage 39:156–168
13. Oztamur D, Karakadilar IS (2014) Exploring the role of social media for SMEs: as a new marketing strategy tool for the firm performance perspective. Procedia-Soc Behav Sci 150:511–520
14. Chen X (2014) Mining social media data for understanding students' learning experiences. IEEE Trans Learn Technol 7(3):246–259
15. Alessa A, Faezipour M (2018) A review of influenza detection and prediction through social networking sites. Theoret Biol Med Model 15(1):1–27
16. Quin L, Sun Q, Wang Y, Wu KF, Chen M, Shia BC, Wu SY (2020) Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Environ Res Pub Health 17:2365
17. Gu Y, Qian Z, Chen F (2016) From twitter to detector: real time traffic incident detection using social media data. Transp Res Part C Emerg Technol 67:321–342
18. Stern EK (2017) Crisis management, social media, and smart devices. In: Akhgar B, Staniforth A, Waddington D (eds) Transactions on computational science and computational intelligence. Springer, Cham, pp 21–33

19. Data age 2025: The digitization of the world from edge to core, https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf. Last accessed 16 2020
20. Hariri RH, Fredericks EM, Bowers KM (2019) Uncertainty in big data analytics: survey, opportunities, and challenges. J Big Data 6:1–6
21. Misirlis N, Vlachopoulou M (2018) Social media metrics and analytics in marketing-S3M: a mapping literature review. Int J Inf Manage 38(1):270–276
22. Wu X, Zhu X, Wu GQ, Ding W (2013) Data mining with big data. IEEE Trans Knowl Data Eng 26(1):97–107
23. Balan S, Rege J (2017) Mining for social media: usage patterns of small businesses. Bus Syst Res 8(1):43–50
24. The Pros and Cons of Social Media in Healthcare, https://etactics.com/blog/pros-and-cons-of-social-media-in-healthcare. Last accessed 20 2020
25. Dai H, Halo J (2017) Mining social media data for opinion polarities about electronic cigarettes. Tobacco Control 26(2):43–50
26. Yang YT, Horneffer M, Dilisio N (2013) Mining social media and web searches for disease detection. J Pub Health Res 2(1):17
27. Corley CD, Cook DJ, Mikler AR, Singh KP (2010) Text and structural data mining of influenza mentions in web and social media. Int J Environ Res Pub Health 7:596–615
28. The Challenges of Social Media in Healthcare and How to Solve Them, https://www.trajectory4brands.com/blog/how-to-solve-the-challenges-of-social-media-in-healthcare-trajectory/. Last accessed 22 Sept 2020
29. Googling health symptoms is problematic, but many Americans still do it, https://www.phillyvoice.com/new-survey-finds-googling-symptoms-problematic/. Last accessed 17 2020
30. How Social Media is Reshaping Todays Education System, https://csic.georgetown.edu/magazine/social-media-reshaping-todays-education-system/. Last accessed 19 2020
31. Wertalik D, Wright LT (2017) Social media and building a connected college. Cogent Bus Manage 4(1):1320836
32. Talaue GM, AliSaad A, AlRushaidan AN, AlHugail A, AlFahhad S (2018) Expected impact of social media on academic performance of selected college students. Int J Adv Inf Technol 8:27–35
33. Faizi R, El Afia A, Chiheb R (2013) Exploring the potential benefits of using social media in education. Int J Eng Pedagogy 3(4):2192–4880
34. How Using Social Media Affects Teenagers, https://childmind.org/article/how-using-social-media-affects-teenagers/. Last accessed 25 2020
35. Chugh R, Ruhi U (2017) Social media in higher education: a literature review of Facebook. Educ Inf Technol 23:605–616
36. Liu W, Luo X, Xuan J, Xu Z, Jiang D (2016) Cognitive memory inspired sentence ordering model. Knowl Based Syst 104(C):1–13
37. Domzouzis K, Akhgar B, Andrews S, Gibson H, Hirsch L (2016) A social media and crowd-sourcing data mining system for crime prevention during and post-crisis situations. J Syst Inf Technol 18(4):364–382
38. National Strategy for Artificial Intelligence, https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf. Last accessed 15 Sept 2020
39. Framework: The Social Media ROI Pyramid, http://web-strategist.com/blog/2010/12/13/framework-the-social-media-roi-pyramid/. Last accessed 26 2020
40. Big Data: A culmination of BI and Predictive analysis, https://www.allerin.com/blog/big-data-a-culmination-of-bi-and-predictive-analysis. Last accessed 28 2020
41. How to Develop a Sense of Presence in Online and F2F Courses with Social Media, https://onlinelearninginsights.wordpress.com/. Last accessed 28 2020
42. Aghabozorgi S, Mahroeian H, Dutt A, Wah TY, Herawan T (2014) An approachable analytical study on big educational data mining. In: International conference on computational science and its applications

# A Cost and Power Analysis of Farmer Using Smart Farming IoT System


**P. Darshini, S. Mohana Kumar, Krishna Prasad, and S. N. Jagadeesha**

**Abstract** Day-by-day India's population is consistently increasing, and it is expected to reach beyond 1.3 billion. In fact after 25 years, serious problems will be predicted regarding food. Hence, it is necessary to develop an innovative system for benefiting the agricultural sector as early as possible. The farmers are suffering due to the reduced rainfall and water scarcity. The traditional farmland irrigation techniques require manual intervention. To overcome this challenge, the proposed research work has developed an automatic irrigation system, which is powered by a smart solar system for saving time, money and work of the farmer. Solar panels will continuously track the position of the sun to ensure the maximum energy production. In addition, an intruder detection system [IDS] is installed into the irrigation system with the help of passive infrared sensor to track where the birds and some other animals are repelled from entering into the field. A Wi-Fi module has been used to establish a communication link between the farmer and the field. The farmer can access the information about the field condition anytime.

**Keywords** Smart farming · Internet of Things · Agriculture · Sensors · Regression analysis · Crop · Estimation theory

## 1 Introduction

Agriculture provided us with a solution to feed all, and it has now come to a halt due to the rapid growth of population. Agricultural lands will not be able to follow the same pace but the worldwide population is expected to grow nearly 10 billion

P. Darshini (✉) · S. Mohana Kumar
Department of Computer Science and Engineering, Bangalore 560054, India
e-mail: mohanks@msirt.edu

K. Prasad
Department of Electronics and Telecommunication, Bangalore 560054, India

Ramaiah Institute of Technology Bangalore, Bangalore 560054, India

S. N. Jagadeesha
PESIT and Management College Shivamogga, Shivamogga, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_21

by 2050 [1]. Agriculture as mentioned earlier in one of the definitions would come under the management of man [5]. With the growth in technology, almost anything can be just controlled with just a click, and in most cases, this is all possible with the help of Internet of Things (IoT). IoT is an ecosystem of connected physical objects. This technology can be incorporated with farming to catch up with the rapid pace of population growth. It would be better to call it smart farming as the entire idea is to make it smart by using technology to increase the productivity by minimizing the cost, power and time as much as possible. A recent Beecham's report entitled toward smart farming agriculture embracing the IoT vision predicts that food production must increase by 70% in the year 2050 [1], in order to meet our estimated world population of 9.6 billion people. It also describes growing concerns about farming in the future: climate change, limited arable land and costs/availability of fossil fuels. Therefore, solution is smart farming. IoT can connect devices embedded, represent digitally, control and connect from anywhere, and then, it helps us capturing more data from more places, ensuring more ways of increasing efficiency and improving safety and IoT security. The main problem farmer faces is that the returns percentage is quite less. This means that farmer returns are not high as much as it should be. This leads to several problems. Many farmers become financially hit due to this, and hence, because of being unable to repay the debts, they end in situations where they end their lives. This is quite common in countries where the global hunger index is high because of the rapid rate of population and the farmer failing to meet the high demands. To overcome this problem, a smart farming system has been developed by using IoT. Smart farming with the help of automation and sensor technology benefits the society such as conservation of water, optimization of energy resources and better crop [14]. The objective of this work is to provide an automatic farming management system which is powered by a smart solar arrangement in this manner saving time, money and power of the farmer. With the automated technology of irrigation, the human intervention can be minimized. Soil parameters like soil moisture, pH and humidity are measured. Then, pressure sensor and sensed values are displayed on the system. If change in parameters such as humidity and temperature of the surroundings, then sensors sense the change and send interrupt signal to the IoT device. With these, farmer will get the alarm notification. The intruder detection system is installed with the help of a passive infrared sensor (PIR sensor) where the birds are repelled from entering into the field. The solar panels continuously track the position of the sun to ensure maximum energy production. Wi-Fi module has been used to establish a communication link between the farmer and field. The farmer can access field condition through the server anytime, hence this likely outcome reducing the manpower and time [2–4, 13].

In the agricultural field, the IoT-based smart farming system predictive and estimation system models play a role, nowadays, to the development of the agro-ecological and socio-economic conditions. In the amounts of resources of the field and farm experiments to provide the information and to identify appropriate and effective management practices. IoT-based model helps to identify the management to cost, time and manpower for land managers and estimation theory approach method use to

make the information save farmers money to great extent and use the smart technique concepts to take the decisions for problem.

## 2 Earlier Work

The following sections present a brief summary on earlier research works carried out in the field of smart agriculture using IoT-based system. Today, IoT leads to the development of many applications such as education, industrial, manufacturing, medical, governance and transportation [6]. The big data predictive analysis techniques are used to analyze crop and cost of fertilizers; IoT technology is used to in the field to collect the data through the sensors and stored in the cloud database through the Internet [7].

R. N. Rao et al. proposed IoT-based smart crop field monitoring and automation irrigation system in the paper discussed [8]. As farmers does not have any thought regarding what amount of water ought to be included and in what manner can the observing of crop should be possible. As huge measure of water is squandered, crop yielding is not appropriate and proposed the accompanying procedures. Brilliant farming is a developing idea, in light of the fact that IoT sensors are equipped for giving data about agriculture fields and afterward follow up on dependent on the client input.

A. J. Raju et al. proposed to build up a savvy agriculture framework that utilizes focal points of bleeding edge innovations, using remote sensor system. Observing ecological conditions is the central point to improve yield of the proficient crops [9].

S Rajeswari et al. proposed new model for crop yield prediction for agriculture field, and it helps to identify the better crop sequence [10].

Wen-Yaw Chung et al. have presented their work in agricultural field combine the cloud and wireless sensor techniques [11, 12]. From the above literature review, a less amount of analysis and data collection were carried out such time, cost and manpower. The proposed system is compressive solution to field activities using smart farming IoT-based technique.

## 3 Smart Farming System

In this work, the smart farming system using the IoT technology is designed in such a way that it is simple to install and use. The proposed system consists of dividing the working of the farm into different areas so that each part can be managed individually. The field consisting of the crops will be installed with the soil moisture sensors, temperature sensors and the PIR sensors. The soil moisture sensor is used to sense the soil moisture and control the drip irrigation correspondingly. The temperature sensor will sense the temperature of the surrounding environment, which will help us in weather analysis. The PIR sensors are installed around the field in specific

locations, which will detect the presence of any intruders, and then, the required action is to be taken. The system includes an efficient water management system to keep a check on the amount of water. Apart from these, the solar system based on the sunflower technology is installed in appropriate part of the field to facilitate the best results. The power system to run proposed system in the field is achieved so by two separate power units. One unit powers the system during the day while the solar system recharges the other, and this second system is utilized to power the system during the night. The main concept of proposed system is that it is smart and automated. The entire system is to be controlled by the microcontroller. The concept of IoT is that it connects all the devices over the Internet. The system is designed in such a way that all the data obtained from the system is sent to a server. Data includes how much water let out, and the soil moisture, temperature, any intruder detection and much more are all sent from the system to the server. The server data used to analyze to make appropriate outcomes. The smart farm consists of a number of components including the field, solar unit and the watering system. In order to achieve the best results, six numbers of prototype model have been constructed in the most efficient way and installed surrounding of Bangalore zone and Mysore zone region India. The farmer can access the system and can keep an eye on all the progressing data being performed in the field. All these are controlled by the Arduino microcontroller. All the sensed values from the field are sent to the microcontroller which then performs the required operations such as controlling solar unit and the drip irrigation system [5].

The entire system is run on the Arduino Uno board which is a microcontroller based on the ATmega328 shown in Fig. 1. The smart farming system has to undergo an entire process of transformation to be able to satisfy the large demands in real world.

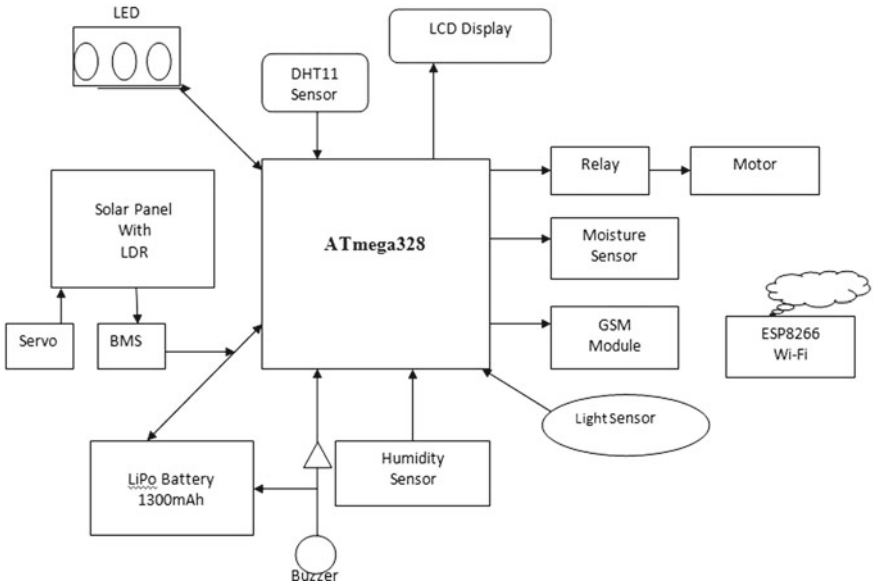



**Fig. 1** Smart farming system

On the whole, the number of devices in the system would increase depending on the size of the field. In the small model, the soil moisture can be tested by examining difference zone region soil moisture. But in real world, since the area of the field is going to be very large, where an efficient method is utilized to install the proposed system, fewer soil moisture can be deployed to sense the soil moisture along an entire crop line. The PIR sensors can be changed based on the area of the field and can be installed in the most appropriate locations so that it can able to identify the intruder on any point on the field. The solar unit which powers the entire system can be scaled up to match the demands of the real-world system.

## 4 Likely Outcomes

To build the system error proof and ensure that it would perform as expected, the proposed system has been tested in many scenarios during the COVID-19 pandemic lockdown periods. The soil moisture sensor is used to sense the amount of water in the soil, and based on the value, it is decided whether the water is fed into the field or not. The capacitance to measure dielectric permittivity of the surrounding medium soil used for proposed system trial works on the value of around 2.9–3.0 V (520 values). If the sensed value goes below this value, then the drip irrigation is activated, and once the value reaches around 650, it stops. Similarly, the PIR sensor is used to sense the presence of any intruder in the field. The sensor has to cover to avoid including the areas of the neighboring fields. The outcomes of the above tests case it is observed that proposed system successively transmit, receive and record the data on server. The data available for transmission through the ESP8266 Wi-Fi module which would utilize the available Wi-Fi connections and exchange the encrypted messages. The data sent consisted of the readings of the moisture, the temperature, number of times water is let out, intruder detected details and much more. Same data also recorded and made to be displayed onto the on field LCD display.

### *4.1 Cost Analysis*

Cost analysis is essential for farm budgeting and planning enabling farmers to effectively compare and determine the profitability of various commodities, thereby creating an opportunity to identify and venture into farming as an enterprise based on current data. The farmer has to consider a number of things when it comes to cost analysis of the farm. By adopting the smart farming system, there are a number of ways in which money can be saved. First and foremost, the entire system runs on solar power. Solar panels are installed that track the position of the sun at every degree. That is, the system consists of a sunflower-based solar system to track the position of the sun, and this is done to ensure that maximum input is captured and converted into energy that powers the system. The solar-based system also recharges the power unit

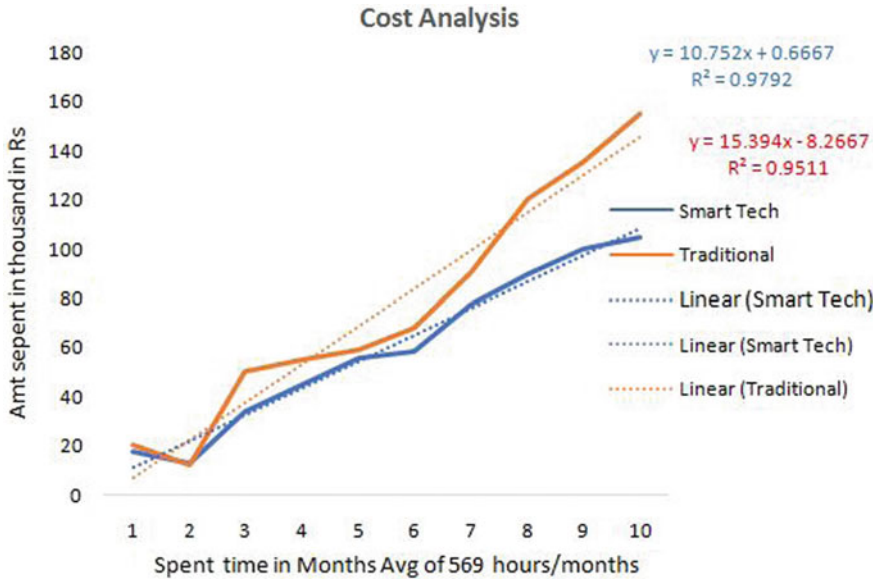



**Fig. 2** Cost analysis

that will be able to run the entire system during night. Based on the farmer feedback and estimation theory approach, this saves the farmers money to great extent. From Fig. 2, a red color indicates the regression analysis from the plotted graph, and blue color indicates the smart farming. This is predictive method using statical modeling regression analysis, and it is clear that expenditure of cost considered parameter is optimally less for small area, as compared to equation plotted in Fig. 2; hence, it has been suggested to compare with the traditional farming, a smart farming is optimally better. This is predictive method able to reduce the power cost by around 10–12% approximately. Another way in which the amount of money can be reduced is by monitoring the amount of water that is being fed into the field. The cost analysis on the data is shown in Fig. 2.

## *4.2 Time Analysis*

A farmer time into field says about its dedication. But this restricts the farmer from being able to perform any of his other activities. Hence, strive to make the system to be able to perform some tasks on its own such that the farmer is able to get save time for himself. This is where our smart farming system into picture. Some important tasks are automated in order to save more time. First, the drip irrigation method is automated by including sensors that sense the soil moisture and let water into the field as and when needed. The amount of time saved here is less but it cannot be

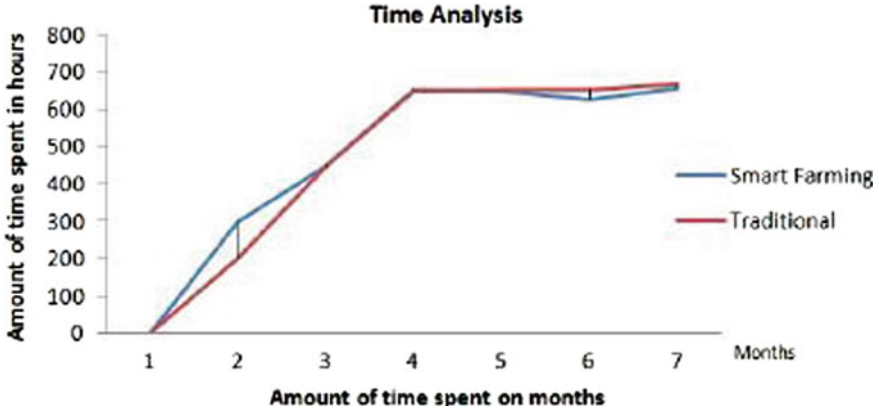



**Fig. 3** Time analysis

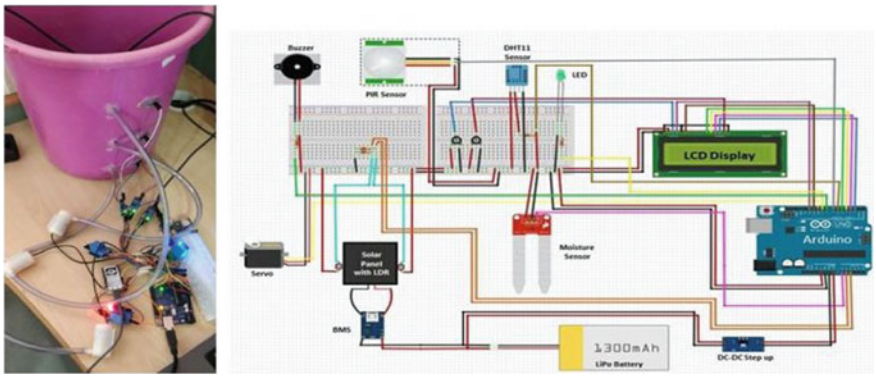
ignored. Some other ways in which time can be saved are by the intruder detection and avoidance system included in our system. Based on the above analysis, the time can be saved by around 8% per annum. The time analysis on the initial data is shown in Fig. 3.

### *4.3 Power Analysis*

Power analysis power not only refers to power to run the system but also the manpower involved in farming. Analysis of power for the system has been explained in the cost analysis section. In the next section tells about more on manpower how can be managed efficiently. A farmer alone cannot run an entire farm. He needs manpower that will work for him on the field and also perform many other tasks. In proposed cost and time analysis, smart farming system with IoT technology automates some of vital tasks of farming. The smart drip irrigation system saves time and cost, but it saves manpower to some extent. The rate at which manpower is saved can only be noticed at the yearly level. The more important aspect is that proposed system would open more career opportunities for people into the field of agriculture to perform tasks such as data analytics, field engineer and may be some more.

## 5 Experimental Setup

The experimental setup is shown in Fig. 4 in two places; the proposed system has been tested in many scenarios during the covid-19 pandemic, where Fig. 4a shows the experiment model, and Fig. 4b shows the schematic diagram of proposed model.




(a) Field Experiment Setup model

(b) Schematic diagram

**Fig. 4** Proposed system model

The soil moisture sensor is used to sense the amount of water in the soil, and based on the value, it is decided whether the water is fed into the field or not. Volumetric soil water content 0–10% for dry soil moisture condition, 11–20% Idel condition.

Water content is measured in two places by isolating the mass of water from the dry of soil is multiplied by 100. The moisture in the soil is characterized as the ratio of the water mass per unit of soil, to the mass of dry soil per unit, and it is expressed as:

$$\mathrm{WC} = \frac{w1 - w2}{w2} * 100 \tag{1}$$

where $w1$ is wet soil + container, and $w2$ is dry soil + container. WC is water content in percentage. Table 1 shows the equation for determining slopes in various sensors. In which $x$ value presents the soil moisture content in percentage, whereas $y$ is cost value is calculated from the equation, respectively.

The proposed system trial works on values of 500–550 values. If the sensed value goes below this value, then the drip irrigation is activated, and once the value reaches around 650, it stops. Similarly, the PIR sensor is used to sense the presence of any intruder in the field. The surrounding area of 120 m is tested in order to sense the PIR sensor. The sensor has to cover to avoid including the areas of the neighboring fields.

**Table 1** Linear equation table and for slope and $R^2$ values

| S. no. | Analysis | Traditional method | Smart tech | Sensor values |
|---|---|---|---|---|
| 1 | Cost analysis | $y = 15.394x - 8.2667$ $R^2 = 0.9511$ | $y = 10.752x + 0.6667$ $R^2 = 0.9792$ | 508…550 |
| 2 | Time analysis | $y = 91.71x + 54.19$ $R^2 = 0.7667$ | $y = 86.6x + 72.44$ $R^2 = 0.786$ | 545…600 |

## 6 Conclusion and Future Work

All experimental and observations test cases prove that and time analysis and the proposed system are a comprehensive solution to field activities, such as smart irrigation and smart warehouse management system, respectively. This approach helps the farmer to improve the yield of the crops and overall production. It is possible to add many additional features to the system based on the demands. An extra water management system can be used to the monitor the water levels, and it can be optimized to get the most efficient results. In the future, farm can also be integrated with smart fertilizer and pesticide feeding system, which would run on satisfying certain constraints. Drones can be included to monitor the entire farm and report to the server so that the farmer can monitor the farm from his home itself can be future work. Data analytics can be improved and can be made to produce better outputs, so that the farmer can be prepared for future. Smart agriculture is already becoming more commonplace among farmers, and high-tech farming is quickly becoming the standard sensors. Future work may be Drone-y estimation that drones can spray fertilizer more times faster than doing so by hand.

*Notes and Comments*. To build the system error proof and ensure that it would perform as expected, the proposed system has been tested in many scenarios during the COVID-19 pandemic lockdown periods at different district regions surrounding Bangalore and Mysore cites. Based on the farmer feedback, the cost and time analysis are presented. The proposed research work was installed, and test beds are started from January to June 2020.

## References

1. Mattos AW (2016) Smart farming, Ph.D. dissertation, Universidadede Lisboa, Faculdade de Arquitetura
2. Amandeep, Bhattacharjee A, Das P, Basu D, Roy S, Ghosh S, Saha S, Pain S, Dey SR, Rana TK (2017) Smart farming using IoT. In: 2017 8th IEEE annual information technology, electronics and mobile communication conference (IEMCON), pp 278–280
3. Mukherji SV, Sinha R, Basak S, Kar SP (2019) Smart agriculture using internet of things and MQTT protocol. In: 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, pp 14–16
4. Oppitz M, Tomsu P (2018) Internet of things. Inventing the cloud century. Springer, Berlin, pp 435–469
5. Suebsombut P, Sekhari A, Sureepong P, Ueasangkomsate P, Bouras A (2017) The using of bibliometric analysis to classify trends and future directions on smart farm. In: 2017 international conference on digital arts, media and technology (ICDAMT). IEEE, pp 136–141
6. Vemuri SR, Satyanarayana N, Prasanna VL. Generic integrated secured WSN-cloud computing U-life care. Int J Eng Sci Adv Technol 2(4):897–907
7. Mohamed F, Osman G, Suhaidi H (2011) Trust management in cloud computing: a critical review. Int J Adv ICT Emerg Reg
8. Rao RN, Sridhar B (2018) IoT based smart crop-field monitoring and automation irrigation system. In: 2018 2nd international conference on inventive systems and control (ICISC), Coimbatore, 2018, pp 478–483

9. Rau AJ, Sankar J, Mohan AR, Das Krishna D Mathew J (2017) IoT based smart irrigation system and nutrient detection with disease analysis. In: 2017 IEEE region 10 symposium (TENSYMP), Cochin, 2017
10. Rajeswari S, Suthendran K, Rajakumar K (2007) A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. In: 2017 international conference on intelligent computing and control (I2C2), Coimbatore, 2017, pp 1–5
11. Chung W-Y, Yu P-S, Huang C-J (2013) Cloud computing system based on wireless sensor network. In: Federated conference on computer science and information systems, 8–11, pp 877–880
12. Srimathi C, Park S-H, Rajesh N (2013) Proposed framework for underwater sensor cloud for environmental monitoring. In: 5th international conference on ubiquitous and future networks, 2–5, 2013, pp 104–109
13. Patil VC, A1-Gaadi KA, Biradar DP, Rangaswamy M (2012) Internet of Things (IoT) and cloud computing for agriculture: an overview. In: Proceedings of Aipa 2012
14. Srisruthi S, Swarna N, Ros GM, Elizabeth E (2016) Sustainable agriculture using eco-friendly and energy efficient sensor technology. In: 2016 IEEE international conference on recent trends in electronics, information communication technology (RTEICT). https://doi.org/10.1109/rteict.2016.7808070
15. Arduino Pro Mini [online]. https://www.arduino.cc/en/Main/ArduinoBoardProMini

# Intelligent Computing Application for Cloud Enhancing Healthcare Services


**Anandakumar Haldorai and Arulmurugan Ramu**

**Abstract** Intelligent computing is a novel means of delivering computing services and resources. Due to massive demand of healthcare services that are transforming the face of health information systems (HISs), investing in healthcare research becomes more fundamental. Nonetheless, just like any form of healthcare advancement, intelligent computing needs to be investigated and evaluated before health practitioners decide to apply it globally. The application of innovations as a result of intelligent computing can be one of the best means of transforming healthcare provisions especially during the process of sharing patient data between medical practitioners during the urgent actual-time cases. However, before initiating the complete transformation of the healthcare sector, there should be a specific strategy that requires thorough evaluation. For instance, a feasible intelligent strategy that is applicable in the healthcare facility must use a public domain cloud infrastructure to permit public accessibility to fundamental engineering healthcare data during the process of retrieving medical resources.

**Keywords** Information technology · Intelligent computing · Cloud infrastructure · Health information systems (HISs) · Electronic health records (EHRs)

## 1 Introduction

Healthcare clinics and public hospitals are in the best position of using publicly available cloud services for remote data storage which is only available in medical domains, and with that regard, this research paper evaluates the concept and present position in healthcare sector. It utilizes four fundamental aspects, i.e., technology,

A. Haldorai (✉)
Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

A. Ramu
Presidency University, Yelahanka, Bengaluru, Karnataka, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_22

management, legal, and security. These four critical aspects are utilized to evaluate the key challenges and opportunities of cloud computing framework for one-stop strategy implementation and procedure which can be utilized by the healthcare sectors to evaluate its strategy, resources, and directions that are used by medical practitioners to shift or transform their services from traditional to cloud-centered healthcare services. This form of transition is also discussed in this literature analysis. Intelligent computing represents the present demand in self-services of the network infrastructure which allows potential users to get access to cloud computing resources in real time from any geographical location to some of the most utilized healthcare applications that include Google docs and Microsoft Hotmail including some of the healthcare-based applications such as Google help platform in the Microsoft Health Vault [1]. Compared to the conventional form of computing, this type of computing has three novel merits: a wide range of computing resources that are present to users, removal of any form of upfront applications by clients, and any form of payment for usage as required. Some literature texts, blogs, and forums have also evaluated the merits and advantages of this form of computing in transportation business, national security, and education.

Many healthcare facilities have already transformed the most used legacy frameworks to incorporate the electronic health records (EHR) which represent a digitalized format of manuscript. Healthcare record systems incredibly follow the standards published by the HIS organization for the clinical healthcare sectors. This form of transformation in the healthcare sector has facilitated the growth of medical provision by nurses, physicians, and administration personnel who have the capabilities of accessing healthcare records whenever required. EHR in most healthcare facilities are presently enclosed in ancient clan's server architectural frameworks. Information technology has incredibly assisted in the process of simplifying healthcare operations as discussed in this paper. This has in turn made it possible for healthcare practitioners to be more effective and client-centric compared to how services were provided a few decades ago.

Embracing the idea of cloud or intelligent computing remedies, it helps to facilitate the smooth running of hospital operations in a convenient and affordable manner. Intelligent computing provides on-demand computing services since it makes use of the latest form of technology in the process of deploying, accessing, and utilizing Internet applications, resources, and information. However, this form of technology incorporates complex system infrastructure that might be problematic to comprehend. In most instances, it is believed that intelligent competing is the most effective selection for learning healthcare activities. This is probably due to the fact that intelligent computing is affordable compared to the cost of installing multiple computer systems in multiple medical facilities.

In case, multiple PC systems were to be installed in every medical room; this system will require efficient software, hardware, networking adaptability, and accessibility to effectively retrieve, store, and upload medical or patient data. For a long time, information technology for healthcare issues has been considered to be relevant in the healthcare sector [2]. Since more sophisticated security and privacy systems have been incorporated in the intelligent computing frameworks, the providers of

services, careers and healthcare industry can professionally identify the control methodologies for specific sensitive patient information. The expenditure of information technology data is on the rise. Cloud-centered electronic healthcare records are constantly having a significant effect on the healthcare sector.

Just like any forms of service operations, the healthcare sector requires more systematic and continuous advancements to remain efficient, timely, and unaffordable to consumers who have a privilege of enjoying high-quality services. A lot of experts and managers can project that intelligent computing is effective for enhancing healthcare services, enhancing healthcare research, and transforming the status of information technology globally. Due to the advent of technology, intelligent computing has significantly reduced EHR start-up costs such as software, hardware, personnel, electronic, and licensing costs that have significantly facilitated the adoption of healthcare technology. According to the research done by researchers, the biomedical informatics society has been able to share applications and data based on the novel computing paradigm.

Apart from that information-handling issue, complexities and unavailable computing remedies are some of the issues that researchers have to focus in the process of evaluating biomedical research information management and evaluation. A number of informatics advancements have also been demonstrated to show that intelligent computing is effective in the process of mitigating these problems. The lack of privacy and security are some of the two major revealing concerns that healthcare practitioners face in selecting the best cloud computing remedy. To effectively mitigate these problems, the healthcare sector should select the most reliable cloud computing provider with the capacity to act according to the provisions and guidelines set by the Health Insurance Portability and Accountability Act (published in 1996).

Due to the constantly increasing breach of big data, there has been a growing uneasiness among healthcare patients who are afraid that doctors and hospitals who utilize intelligent cloud services will complicate the security and privacy of their personal data. Apart from that, there still is a prevailing issue of permitting multiple users to use and share EHR throughout the healthcare facility. In addition to the patient safety and privacy information, breaches are expensive to deal with since the healthcare sectors over the past few decades reported to have lost millions of funds annually.

According to the recent research on the fans of data breaches lost, analysts from Pennamon School argue that healthcare quote losses have been reported to be twice as much as the global average which is typically $380 in 2017, where the global average is reported to be $141 [3]. Irrespective of their wide-range benefits connected to diligent computing applications for the medical sector, there are several technological, management, legal, and security issues that should be mitigated. In that regard, this paper focuses on discussing the aspect of intelligent computing, its present application in the medical field, its opportunities, its challenges, and how the concept can be implemented for strategic planning during the introduction of new models of services in the healthcare facility.

## 2 Literature Review

According to many previous studies, the adoption and status of intelligent computing in the healthcare system have been evaluated. These literature texts explain the potential advantages of intelligent computing, the projected frameworks, and models that are supposed to enhance healthcare provisions. Among some of these projections, researchers have conducted an analysis on the cloud-centered framework that is meant to automate the process of gathering patients data using a system of sensors linked to the legacy medical devices. Moreover the system is vital for the processing and delivering datasets to healthcare centers which have to be stored, processed, and distributed. The main advantage of healthcare systems is to provide users with daily and actual-time datasets in an automated manner, hence eliminating manual gathering of data and typing including deployment of healthcare services.

Researchers [4] have also evaluated cloud-based computing standards for management systems to effectively provide multimedia sensing signals and security frameworks as vital services for users of mobile devices. Apart from that, the system has also relieved mobile devices from undertaking mass multimedia and privacy algorithms that facilitates the process of delivering mobile healthcare services. This will therefore enhance the application of intelligent mobile devices that are essential for community services which also enhance healthcare provision to marginalized and vulnerable communities. In another research done by researchers, aversive cloud initiatives have also leveraged the impact of intelligent computing and wireless advancements in the process of allowing physicians to retrieve vital health data at an actual time from any geographical location. Researchers have also explained the prototype emergency healthcare framework for the Greek national healthcare service enabling it to integrate emergency frameworks with individual healthcare records insistence that is meant to provide healthcare practitioners an immediate accessibility to patient personal data from any geographical location and the use of any computing device at an affordable price.

A wide range of resources and articles has also evaluated the effective application of intelligent computing in bioinformatics analysis. For instance, researchers have projected the system-based computing framework and concept that is applicable in the colorectal cancer analysis meant to alouette critical imaging in clinical operations. Due to this research, the application of the Amazon EC2 services that include 100 notes has been applied in the process of assembling a complete human genome approximately 140 million individuals who are capable of reading sequences from the sequence searches and alignment of hashing algorithms. Researchers [5] also appreciated the application of the Amazon EC2 in the process of computing autologous connections of more than two hundred thousand in numbers which have also been compared in multiple research analysis. This kind of computation took the researchers more than 200 h and a fund of $8000 to complete this research. The analysis also applied intelligent computing in the process of evaluating the implication of the G Quadruplex onto the Affymetrix array.

The laboratory system for individualized medical activities and biomedical informatics has also appreciated the advantages of intelligent computing in the process of developing genetic science in economics that will potentially manage massive amounts of data that are recorded in real time. Other than this academic analysis, most globally recognized software industries have significantly invested in cloud computing hands extending their novel services and offers for healthcare records services such as the Microsoft HealthVault, the Amazon's Web service, and the Oracle exalogic elastic cloud system which represents some of the services and technologies used in the storage of individual healthcare data in an online system. Moreover, application of the healthcare intelligent computing system has been appreciated by researchers globally. For instance, Amazon's Web service application allows providers and healthcare practitioners to gather HIS offerings that are considered a scalable storage system infrastructure in a healthcare facility.

According to the American occupational network, improvement of patient care through the application of modernized healthcare record systems is essential to facilitate the process of updating healthcare processes under a cloud best software application. Many companies at the moment can avail accurate and prompt services in real time, hence minimizing any medical transcription expenses by approximately 80% [6]. The department of healthcare and human services in United States is in conjunction to the national coordinator for healthcare information technologies that have recently selected acumen remedies and intelligent customer relationship management systems to effectively select and implement EHR frameworks through the country.

This software is effective since it allows localized extension facilities to effectively manage and interact with medical practitioners on matters concerned with the application of the EHR framework. The Royal College in Australia has gathered healthcare practitioners to work together in the process of developing an electronic healthcare cloud system with the assistance of a leading telecommunication provider Telstra in Australia. The medical school is composed of more than 20,000 members who have the access to sophisticated medical facilities to build a modernized electronic healthcare facility. The software is advantageous in a number of ways: Developing medical software, introducing care plans, facilitating quality medical training, making proper medical decisions, and enhancing proper management services in the healthcare facility.

## 3 Intelligent Computing: A Novel Economic Computing Framework

Cloud computing is still a constantly enhancing development whereby its attributes, definitions, and features are deemed to advance in the next few years. However, there are a lot of definitions that have been proposed over the past few decades including the mini mall explanation that incorporates fundamental features of the concept. In

relation to this paper, a cloud system is described as a massive segment of easily accessible and usable virtual resources, i.e., development platforms development services and hardware systems. All these resources can vibrantly be recognized by practitioners to effectively adjust to the variable scales which have the capacity to optimize the utilization of resources. This massive segment is normally exploited on the basis of a pay-per-usage framework which assures providers of an effective infrastructure that delivers expected results in real time.

Considered from the perspective of a typical service, intelligent computing incorporates three essential archetypal frameworks: platforms, infrastructure, and software.

- Software-as-a-service (SaaS): Some applications such as HR are normally controlled in a cloud service provider before being unveiled to two specific customers through a specific network connection, i.e., the Internet.
- Platform-as-a-service (PaaS): This enhancement application, i.e., operation framework, is typically hosted in an intelligent computing system which can only be accessed through the Internet with PaaS, and healthcare practitioners can work with developers to build structural Web applications without necessarily installing any form of tools in their PC systems before deploying them to users without any form of administrative knowledge.
- Infrastructure-as-a-service (IaaS): With this application, the cloud users can effectively outsource useful tools for supporting clinical operations with the assistance of storage systems, hardware, applications, and networking elements. The application programmers have the capacity to possess the tools and assign users the responsibility of maintaining, running, and housing these applications. However, the users are expected to enjoy the application on a pay-per-use basis.

To effectively deploy intelligent computing, the United States National Institute of Standards and Technology (NIST) has recommended four essential frameworks.

- The public cloud: Cloud service providers have to make resources such as storage and application systems to be available to users in real time via a networking system. For instance, the Amazon's elastic computing cloud system (EC2) permits its users to utilize virtual computers on a rental basis which also allows operators to utilize their own customized applications. EC2 operates within the Amazon network data centers and data infrastructure that allows users enjoy the services at a minimal fee.
- The private cloud: Intelligent cloud infrastructure, over the past few decades, has been used by many organizations for sharing a single concern which include privacy requirements, security, mission accomplishment, and standard requirements. For instance, the Microsoft azure permits clients to establish foundation for encrypted cloud infrastructure by using the windows service and the system centers for creating growth products with dynamic information center toolkit.
- Community cloud: The cloud infrastructure represents the private, community, and public domains of the cloud system. Here, the healthcare sector is capable of managing and providing privacy requirements, policies, missions, and compliance

standards. The communication cloud-enabled the Los Angeles City council to make use of the Google cloud system to safeguard its data and applications that can only be retrieved and accessed by the agencies working for the city council.

- The hybrid cloud system: The hybrid cloud infrastructure incorporates the cloud system and community system. This system allows an organization to manage and provide IT services within its own domain which fails in certain services in an external domain. For instance, IBM can effectively collaborate with the Junipers network system to avail hybrid cloud services and infrastructure to companies that need to advance their private cloud systems at remote areas in an accessible manner.

## 4 Critical Analysis of the Healthcare Cloud Computing Challenges and Opportunities

According to recent research, approximately 75% of the basic information and datasets from office hours have reported the necessity of applying intelligent computing in the process of advancing the future status of healthcare services. The projections indicate that the number of users using the mobile cloud system is proposed to increase from 70 million to about 1 billion over a couple of years to come. In the medical sector, a wide range of organizations, experts, and managers have the belief that intelligent computing can incredibly increase a fundamental service required in medical activities [7]. Apart from that an analysis by the European network and the data security agency in the UK stated that the novel computing framework is projected to enhance an incredible global investment in information technology by many companies and industries including the healthcare sector.

The report also projected that by the end of 2020, the USA alone would have spent an approximation of 50 billion on intelligent computing applications which are essential for the advancement of healthcare services. Just like any form of innovation intelligent computing, it has to be evaluated thoroughly before being applied internationally. A number of research analyses have recently evaluated the implication of cloud computing in the medical sectors which have been discussed in this research, in terms of challenges and opportunity from the perspective of technology, management, legality, and security. The subsection below discusses these challenges and opportunities in detail.

# 5 An Evaluation of Cloud Computing Challenges and Opportunities

## *5.1 The Management Perspectives*

**Opportunities**

The key merit of intelligent computing is its affordability. For instance, the Amazon Company charges users about 0.1 dollars every hour for 1.0 GHz instruction data architecture for a typical EC2 service. The company also charges users about 0.12 dollars for 1.0 GB data per month and transfers information across Amazon's Web service network. In that case, the company gets cost-effective advantage, while it is also considered to provide quality information technology solutions through intelligent computing without necessarily purchasing or evaluating hardware and software or requiring users to pay for in-house technological infrastructure. Resultantly, this makes it possible for the company to concentrate on crucial applications without worrying about additional expenses such as training IT staff. Moreover intelligent computing methodologies enhance the deployment of services where it helps in maintaining crucial flexibility which means rapid accessibility and ubiquitous elasticity to healthcare resources. This capacity implies that even though demands of services change, medical facilities and hospitals do not require transformation of their systems and infrastructure as a result of these transformations.

**Challenges**

In intelligent computing, the main issue is about the privacy and security of users in retrieving their personal data, governance loss, management inertia, and uncertainties on dealing with the provision of services. Maintaining trust when using a specific service is a key aspect in business that has to be assured in intelligent computing as well. Challenges normally tend to occur when sensitive information and mission-essential applications are transferred into the cloud system where network providers do not effectively assure the privacy and security of data systems [8]. Organizational inertia also known as cultural resistance in the process of sharing data is capable of transforming the manner in which typical management challenges are mitigated in intelligent computing. In case a provider is not capable of meeting the compliance standards, regulations, applicable protocols, and policy transitions, networking users might have their activities rendered to risk. In most cases, particular service provisions such as money transactions cannot be undertaken in such cases.

## 5.2 Technological Aspects

### Opportunities

Small-ranged hospitals, laboratories, and medical practices normally do not include an internal information technology staff to ensure the maintenance of in-house infrastructural systems for mission-critical applications like an EHR. In that case, dealing with novel infrastructural costs and Internet technology maintenance costs can effectively mitigate a wide range of obstacles following EHR adoption. In case of wide-ranged hospital organizations, incorporating data storage systems and technological applications is essential for the intelligent computing system to effectively manage and mitigate issues from a third-party network provider. Based on the perspective of technological management, intelligent computing can significantly advance the level of flexibility and scalability which resultantly increasing the affordability of computing infrastructural services. Moreover intelligent computing has some merits which can be realized from the green computing that is known for its efficiency and application in the ecosystem to promote energy savings. Application of already available resources established in a healthcare system significantly enables providers to provide cost-effective services. In an energy system for instance, electricity costs can drastically be rationalized to help providers to save the resources required to establish PC systems and other computer components in an organization. This helps to minimize the production of dangerous resources or materials in the ecosystem.

### Challenges

Some of the technical issues related to the utility of intelligent computing include the exhaustion of resources, performance unpredictability, information lockage, and potential bugs at a massive scale in the distributed intelligent system. The minimal costs and intelligent computing resources that are present when demanded by users are two fundamental characteristics of cloud computing. Nonetheless, the healthcare sector is gradually becoming crowded with more technical service providers coming to market. Due to the rapidly increasing demand of services subjecting to competition, a lot of cloud computing providers commit themselves to cloud computing resources such as storage spaces, mobile applications, and CPU locations to attract potential clients.

To effectively manage profits in an organization, the value of delivery system has to be evaluated in a manner that suits the interests of clients. For instance, providers might decide to limit accessibility to cloud computing resources or focus on introducing outdated software or hardware which target on shuffling CPU technology. However, a lot of cloud computing customers are not capable of controlling virtualized system architecture which makes it necessary for network providers to deny permit to audits from potential clients. As such, it might be challenging to predict the performance of a specific healthcare sector. This form of a relation between the expectation of a client and healthcare providers can significantly lead to potential challenges for cloud service clients who are interested in providing high-quality services to end users.

Data lockage is also a fundamental challenge that has to be analyzed critically. In most instances, cloud computing clients might be stimulated to shift their services or data to another cloud provider to enhance the process of establishing an in-house information technology ecosystem in case service providers cease their service or business operations. For instance, Google made a decision to halt its Google Health Services in 2012 which stimulated users to take an approximate of 12 months to retrieve their healthcare information. However, some of the intelligent computing infrastructures provide minimal applications, data, and services to the users.

As a result, it poses significant challenges to clients who focus on migrating from a single service provider to another or transferring the services or data back to the in-house information technology ecosystem. A number of intelligent computing users such as biomedical research laboratory might require frequently downloading or uploading significantly massive amounts of data from the cloud system [9]. The users of mobile applications might have noticed some form of information such as transfer bottleneck as a result of physical network bandwidth limitations. Apart from that users might face the challenge of handling technical risks since bags might have been distributed in the intelligent cloud system in abundance. When these challenges are compared to the in-house information technology system, some of the mistakes or bugs pose a significant challenge in debugging errors in the massive-scale distributed cloud infrastructure.

## *5.3 Privacy Aspects*

**Opportunities**

Actually, one of the most vital concerns in the process of incorporating intelligent cloud computing in healthcare facility is linked to data privacy and security. However, in contrast to the localized housed data system, this framework is not fully secure. In some instances, it normally enhances security and privacy since cloud service providers such as Amazon, Google, and Microsoft are capable of devoting massive amounts of resources to mitigate any privacy and security challenges which might be affordable to users [10]. This might significantly help in mitigating any potential distractions of healthcare legal documents and medical records which is one of the challenges that the New Orleans hurricane disaster posed to the healthcare sector.

All forms of security and privacy measures such as software or hardware, human resource management, and the control of funds are affordable whenever implemented on a massive scale. A number of intelligent cloud service providers have a tendency of replicating customers' data in multiple geographical locations which is an aspect used to enhance the redundancy of data and system independence in case any potential failures of the system are reported. Apart from that, the cloud service providers typically have the capacity to vibrantly reallocate privacy resources for the process of data filtering, traffic modeling, and data encryption to enhance any potential support for the available defensive measures such as the denial-of-service attacks on critical

healthcare services. The capacity to vibrantly scale-out defensive measures whenever demanded by users has significant advantages for system resilience.

**Challenges**

In information technology, many data privacy challenges have been reported. These include network breakage, natural phenomena, hackers' intrusion, management failure, poor data encryption, management, and data privilege mismanagement. Certain threats in the intelligent computing sector include privilege misuse, poor data encryption management, and failure to manage public data. Intelligent cloud computing is typically available in multiple geographical locations for mobile users. For instance, if a client requests to eliminate data encrypted in virtual computing infrastructure with sophisticated operating frameworks, this might potentially make data systems to erase fundamental information. The sets of data are available in a disk which can only be accessible by the service providers. For multiple hardware resources and tenancy, ecosystems used by customers' datasets might be accessed or deleted by a third-party provider [11]. As a result, this puts the encrypted data into significant risk for the cloud computing users.

## *5.4 Legal Aspects*

**Opportunities**

Information and security privacy protection is the fundamental for enhancement of client trust required for intelligent computing to access its massive potential. In case the provider adopts clear, better practices or policies, users are in the best position to assess potential risk during the use of the cloud computing system. Advantageously, a lot of the main providers are committed to enhance and formulate the best policies and user practices to safeguard clients' privacy and data. Other than the commitment of these providers, a number of computing organizations like the cloud security alliance have formulated an incredible and well-understood guide to handle privacy and security problems. Such organizations have also structured software and hardware initiatives which are meant to establish the most secure platforms for cloud computing. The Federal government is also essential when it comes to fostering globalized agreements and regulations for both providers and the users.

**Challenges**

The application of intelligent computing presents a wide range legality challenges like intellectual property rights, contract laws, information jurisdiction, and privacy rights. Among them, information jurisdiction and privacy challenges are some of the major prevailing issues. In the cloud computing sector, the physical storage is distributed widely over a lot of jurisdictions whereby each of them is governed by a set of laws that are designated to usage, privacy, and intellectual user's rights and properties. For instance, the health insurance portability and accountability act in the

USA has placed restrictions on companies that disclose individual health information to third parties who are not affiliated to the healthcare sector. In that case, these regulations provide the Federal government mandate to retrieve and ask for data in case conditions are deemed, essential, and emergent to the homecare security framework. As such, these acts potentially regulate the power of industries and organizations to utilize, disclose, and collect personal data for commercial purposes. Nonetheless, providers without the consent of users might transfer the data from one direction to another. Information in the cloud system is composed of a single legal location at a specific time with variant legal challenges.

## 6 Discussion

### *6.1 Intelligent Computing Strategy and Planning*

Whenever a healthcare organization considers shifting its services into a cloud system, it requires strategies and planning frameworks to analyze its novel benefits, potential risks, and also evaluating capacities to accomplish missions and establish the planning frameworks scheduled for implementation. There are a lot of references for the creation and establishment of a secure intelligent computing strategy and planning framework. For instance, organizations can use the cloud computing life cycle that has been adopted to include nine stages to assist users to launch intelligent computing projects. These are some of the proofs, concepts, piloted projects, and road maps that enhance modeling, integration, maturity, collaboration, expansion, and planning of the computing system. The institute of project management which is a non-profit membership organization for project management professionals did the research on intelligent computing which can be used by cloud project managers. This research provided eight fundamental steps that are required in the implementation of intelligent computing.

The federal health information technology strategic planning that was established by the Federal government in 2008 can be significantly used by large government in detail for the process of implementing healthcare cloud computing projects [12]. The strategy and planning framework provided a mandate to the national coordinator for health formation technology to enhance the development of local implementation of the HIS infrastructure. This duty is purposed to enhance efficiency and quality in the healthcare sector. The strategic planning is also projected to have some fundamental merits: patient-concentrated health with a purpose to enhance privacy and security, adoption, interoperability, and associative government.

The projections to attain every objective have been proposed in approximately 43 strategies which illustrate the works required in attaining every merit and objective. Every planning framework is connected to the major milestone over which the developments might be evaluated, and some actions are undertaken to effectively implement certain strategies. Other than the discussed strategic methodologies, this

research considered the recommendation of healthcare cloud computing planning frameworks known as the $HC^2SP$ model which might be utilized by the healthcare sector for cloud-centered services. This framework incorporates four fundamental stages: identifications, evaluations, actions, and follow-ups.

**1st Stage: Identifications**

In the $HC^2SP$ model, the initial stage is meant to evaluate the present condition of the healthcare sector; its process is to effectively identify fundamental objectives to critically enhance the improvement of services which utilizes the voices of patients and customers. The major evaluation methodologies might be essential to evaluate the issues of the present service processes. The process of patients' admission in the healthcare facility is considered too long. There is significantly unessential charting throughout the process due to the absence of automatic data systems like electronic medical records (EMR) and HER which incorporate a lot of upfront information technology maintenance and investments. The objective identifications and its entire scope might be clarified to effectively serve the entire end users who are the patients in this case. Moreover, strategic planning groups have to provide an illustration of the healthcare services including an explanation of the healthcare quality indicators and their main obligations. In this stage, the team is able to structure a defined scope of service issue that the facility faces.

**2nd Stage: Evaluations**

In the second stage of the framework, an evaluation of the potential challenges of launching cloud computing and intelligent security alliance is done. NIST have created a comprehensive standard which is meant to analyze the advantages and challenges of implementing intelligent computing with this potential users might be able to apply the SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis in the process of evaluating the feasibility levels of intelligent computing or the cloud-centered approaches. Moreover, users have to access the methodologies that are meant to deal with the identified problems. In the present literature, there are some of the analyses that evaluate the purpose. For instance, researches have reported a number of obstacles following the implementation of the intelligent computing. Every obstacle has been computed to have some potential solution and opportunities which is varied from product management to fundamental projects. Researches have also evaluated a number of internet-federated organization challenges like the process of information transfer bottleneck, sharing logs, and the federated distributed cluster.

Apart from that, these researchers have also recommended further research evaluations meant to mitigate the discussed challenges. Moreover, researches have recommended the XML-centered mediators meant to deal with data locked issues. The intelligent cloud-based security alliance illustrates about twelve domains that are linked to cloud computing. These are some of the domains that are categorized into two significant segments: operation and governance. Remedies and recommendations might also be categorized in every domain [13]. According to the NIST standards about privacy and security of the publicized cloud systems, the privacy issues and security frameworks should correspond to the precautions recommended by the

healthcare facility. These standards and guidelines have to be followed whenever initiating and planning publicized cloud services.

**3rd Stage: Action**

Once novel computing frameworks have been evaluated, the organization might be capable of determining whether new services have to be adopted or not. If that is the case, there is need to upgrade an implementation framework. In this process, there are five essential steps that should be considered.

Firstly, the organization has to determine the intelligent computing service models. As evaluated in this paper, cloud computing represents a number of various services (IaaS, PaaS, and SaaS), including various models of deployment such as hybrid, community, public, and private cloud). Every service model is defined by its individual risks and benefits. In that case, the major consideration in differentiating the forms of deployment and services should be varied.

Secondly, there is need to compare the various cloud service providers. Selecting the most effective cloud provider is the most essential segment of a proper implementation plan. Various providers might project various frameworks, audit procedures, pricing frameworks, security, and privacy guidelines. As such, the healthcare organizations should facilitate the comparison of various offerings. Moreover, there is need to evaluate the providers' performance and reputation before signing any form of contract.

Thirdly, obtaining assurance from the chosen cloud computing providers should be facilitated. The organization has to accept assurances which the chosen providers might use to avail the quality of services (QoS), imitate sound security, privacy, legal regulations, and practices. The QoS assurance incorporates the pay-per-use, on-demand accessibility, timely troubleshooting, and operation transparency. The assurance of security and privacy handles the information availability, integrity, confidentiality, non-repudiation, and authorization. Moreover, the providers have to assure that information and the essential backups are encrypted in the specified geographical locations through service management frameworks, service regulations, and contracts [14].

Fourthly, there is need to consider future information migration. The healthcare sector may be required to shift services and data to other service providers or essentially back it up in a form of an in-house IT ecosystem since the providers might cease their service or business operations [15]. For instance, the recent ceasing of the Google Health followed the diminishing of services' quality which has also amounted to the disputing of contracts. The portability of data has to be viewed upfront as a portion of this plan [16].

Fifthly, it is essential to consider pilot implementation in the healthcare sector. A lot of previously implemented planning aspects suggest that organization without any cloud computing experience might consider pilot implementation. The process of piloting has to be suitable to effectively facilitate the process of cloud computing in the healthcare sector [17].

**4th Stage: Follow-ups**

Initiating follow-ups is the final segment in the process of deploying intelligent cloud computing infrastructure which is based on a proper follow-up plan. The structured plan shows the points when the organization has to measure or the manner in which this measurement has to be done for the purpose of improving potential services. Considerable targets have to be established prior to the analysis of results when novel services are valuated over the specific targets or whenever the performance evaluators have to be assessed based on the magnitude of service improvement.

## 7 Conclusion and Future Directions

In conclusion, cloud computing is a novel framework which assures affordability, efficiency, and flexibility in the information technology services to the end users. Moreover, it provides the chance for enhancing the adoption of HER, research services, and the healthcare services. Nonetheless, as evaluated in this paper, there are some potential issues in the process of fostering novel frameworks in the healthcare sector. Certainly, the most firm resistance to launch cloud computing in the HIS segment is connected to legacy issues and information security. However, a lot of the main providers such as Amazon, Google, and Microsoft have focused in the process of developing the best practices and policies to effectively secure the clients' privacy and data. A number of the non-profit organizations like the cloud privacy alliance and the entrusted computing segment have effectively developed more comprehensive software and hardware technologies, which effectively enhances the construction of cloud computing applications. Apart from that, the federal government has also fostered some regulations such as PIPEDA and HIPAA to safeguard cloud users information privacy and security. Moreover, a number of legal regulations in the intelligent cloud computing typically revolve in the contract negotiation and evaluation. In future, providers should put more focus on the development of $HC^2SP$ that is applicable in the healthcare sector to effectively determine its strategy, resources, and directions to shift to the paradigm of cloud computing. The framework incorporates four fundamental stages such as follow-ups, action, evaluation, and process identification. The future works should focus on further expanding the SWOT evaluation which will definitely allow organizations to evaluate and determine potential of launching the novel model, and also some of the remedies to deal with cloud computing issues also require more in-depth research.

## References

1. Kobashi S (2011) Guest editors. Int J Intell Comput Med Sci Image Process 4(2):89–91. https://doi.org/10.1080/1931308x.2008.10644185

2. Hata Y (2007) Opening editorial. Int J Intell Comput Med Sci Image Process 1(1):1–2. https://doi.org/10.1080/1931308x.2007.10644139
3. Ḥusein A, and Urquhart A (2020) Clinical decision support system for the activity of evidence based computation. J Med Image Comput:50–57
4. Aaron PL, Bonni S (2020) An evaluation of wearable technological advancement in medical practices. J Med Image Comput:58–65
5. Karthikeyan K, Prakash EP (2020) Web based analysis of critical medical care technology. J Med Image Comput:66–73
6. Haldorai A, Anandakumar S (2020) Image segmentation and the projections of graphic centred approaches in medical image processing. J Med Image Comput:74–81
7. Wachs J et al (2008) Real-time hand gesture interface for browsing medical images. Int J Intell Comput Med Sci Image Process 2(1):15–25. https://doi.org/10.1080/1931308x.2008.10644149
8. Kobashi S, Nagamune K (2007) Special issue on medical engineering. Int J Intell Comput Med Sci Image Process 1(2):91–92. https://doi.org/10.1080/1931308x.2007.10644141
9. Hata Y (2013) Special issue on 2nd conference on Himeji initiative in computational medical and health technology, University of Hyogo. Int J Intell Comput Med Image Process 5(1):1–2. https://doi.org/10.1080/1931308x.2013.798989
10. Dessouky M, Elrashidy M, Taha T, Abdelkader H (2014) Computer aided diagnosis system feature extraction of alzheimer disease using MFCC. Int J Intell Comput Med Sci Image Process 6(2):65–78. https://doi.org/10.1080/1931308x.2015.1004823
11. Fujita T et al (2013) Daisy-chain shape wearable health monitoring system by using fuzzy logic heart-rate extraction. Int J Intell Comput Med Sci Image Process 5(2):125–133. https://doi.org/10.1080/1931308x.2013.847621
12. Kobashi S et al (2013) Newborn brain MR image segmentation using deformable surface model based on fuzzy knowledge models. Int J Intell Comput Med Sci Image Process 5(2):115–124. https://doi.org/10.1080/1931308x.2013.847618
13. Uozumi Y et al (2013) A bone segmentation for the knee joint in MDCT image based on anatomical information. Int J Intell Comput Med Sci Image Process 5(2):105–113. https://doi.org/10.1080/1931308x.2013.847617
14. Takeda T, Kuramoto K, Kobashi S, Hata Y (2013) A fuzzy moving object estimation using infrared TOF camera. Int J Intell Comput Med Sci Image Process 5(2):147–160. https://doi.org/10.1080/1931308x.2013.838068
15. Kitamura K et al (2013) Development of salivary NO3−measurement device for navigators' mental workload. Int J Intell Comput Med Sci Image Process 5(2):135–146. https://doi.org/10.1080/1931308x.2013.847636
16. Quintanilla-Domínguez J, Ojeda-Magaña B, Marcano-Cedeño A, Barrón-Adame J, Vega-Corona A, Andina D (2013) Automatic detection of microcalcifications in ROI images based on PFCM and ANN. Int J Intell Comput Med Sci Image Process 5(2):161–174. https://doi.org/10.1080/1931308x.2013.838070
17. Momennezhad A, Ebrahimnezhad H, Shamsi M, Asgharian L (2014) Brain activity EEG-P300 signal categorization from LPC based estimation of signal using fisher linear discriminant analysis. Int J Intell Comput Med Sci Image Process 6(1):17–26. https://doi.org/10.1080/1931308x.2014.925271

# Coronavirus Detection and Classification Using X-Rays and CT Scans with Machine Learning Techniques


**Moulana Mohammed, P. V. V. S. Srinivas, Veldi Pream Sai Gowtham, Adapa V. Krishna Raghavendra, and Garapati Khyathi Lahari**

**Abstract** This work aims to detect the signs of the Coronavirus, also known as Covid-19. A dry cough, sore throat, and fever are the most common symptoms of Covid-19. For Covid-19, it is important to find fast and precise results at the time to stop it in the early stages and to avoid it from the vast spread. To interpret and identify the symptoms from X-rays and CT scan images, the machine learning and computer vision principles were applied. The works are usually performed with the CSV datasets. However, the analysis is performed to compare the images of patients with Covid and Non-Covid. To enhance the classification performance, it is feasible to use feature extraction techniques such as CNN, local directional pattern (LDP), gray-level run length matrix (GLRLM), gray-level scale zone matrix (GLSZM), and discrete wavelet transform (DWT) algorithms (Barstugan in Coronavirus (Covid-19) classification using CT images by machine learning methods [1]). In this paper, the convolution neural network model is selected as the classifier. Softmax is used during the classification process to classify the images given, whether they belong to Covid or Non-Covid. This implementation is carried out on both the X-ray images dataset and the dataset of CT scan images which are obtained from the repository that is publicly accessible.

**Keywords** Covid-19 detection · Coronavirus · CNN · CT scans · X-rays · Machine learning · Computer vision

## 1 Introduction

The outbreak of the Covid-19 Coronavirus, SARS-CoV-2, has raised a catastrophic worldwide crisis end of the year 2019. Day by day, the cumulative incidence of Covid-19 is steadily increasing. Cloud computing and machine learning (ML) can be used very efficiently to monitor the outbreak, forecast the epidemic's progress,

M. Mohammed (✉) · P. V. V. S. Srinivas · V. P. S. Gowtham · A. V. Krishna Raghavendra · G. K. Lahari
Department of Computer Science and Engineering, KL University, Vaddeswaram, Guntur, A.P., India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_23

and model strategies and policies to control its spread. To evaluate and forecast the development of the epidemic, this study applies an enhanced mathematical model. Forecasting the potential hazard of Covid-19 in countries worldwide and improved ML-based model has been applied. A better fit to create a prediction system which can be obtained using the iterative weighting to fit the generalized inverse Weibull distribution. This has been implemented with a more precise and real-time prediction of the epidemic's growth activity on a cloud computing platform. For a constructive response from the government and people, a data-driven strategy with greater precision can be quite useful. Finally, for more practical applications, a collection of research opportunities and setting-up grounds were suggested. To make a prediction, it is required to determine which algorithm can be effectively used to classify the given image into a favorable or unfavorable result. This paper aims to work with convolutional neural networks and classifies the category using softmax.

## 2 Literature Review

The expert radiologists detected from CT images that Covid-19 shows different behaviors from other viral pneumonia. Therefore, clinical experts specify that the Covid-19 virus needs to be diagnosed in the early phase [1]. At the beginning of this research work, it is planned to work with a CSV file, but after finding many empty spaces and incomplete datasets, the implementation is done with CT scans and X-ray images [2, 3]. These days when access to the corona data is very limited, experiments are carried out with a dataset with very little data and inter-class imbalance. Due to the limitations of the dataset, feature extraction methods are preferred over deep learning-based methods [2]. The working on images dataset rather than CSV dataset, it is quite difficult to find CT scan and X-ray images datasets. After searching numerous sites and repositories, a public repository is sorted out with both CT scans and X-ray images [4]. After selecting the dataset, the images were not in a similar way to get the proper result. Preprocessing the images of both datasets is using computer vision techniques mentioned in [5], which guided us to convert all images into a single format 'png'. The next task is to convert all the images into the same dimensions for processing, done by verifying the work in this citation [6]. While researching the working with CNN, the best in feature extraction of images was recognized. Referred some articles to know the working of the algorithm and to observe the results of implementations [7–9]. To gain the working and processing knowledge of convolutional networks, some research were undergone to know the exact implementation of this algorithm [10, 11]. After going through all the data, it is decided to implement CNN for feature extraction from images.

## 3 Proposed Methodology

In this paper, the classification is applied with Covid and Non-Covid patients from their CT scans and X-rays. Apart from the normal approach, this paper uses an images dataset which is publicly accessible by every one of CT Scans and X-rays of Covid and Non-Covid patients for best classification. At first, convert all the images into a single format which is 'png'. Later resize the dimensions of every image into 128 $\times$ 128 by reducing the size of the dataset. That resulted in the dataset is used in the classification. Both the Covid and Non-Covid datasets are combined, shuffled, and divided into train and test. The sequential model is used because only the single input and single output were obtained. In this model, three convolution layers and two hidden layers are used. Early stopping is used to stop the loss. The softmax function is used to predict the score and classify the image into either Covid or Non-Covid. The results of implementation are compared in both CT scan and X-ray images (Fig. 1).

## 4 Experiments

### *4.1 Dataset Description*

For better classification and a better outcome, the input of CT scan and X-ray images was directly considered. By this approach, the there will be no missing of features during the implementation of CSV dataset. These CT scan and X-ray images are publicly available. The dataset contains both Covid and Non-Covid patient's lung images of both CT scans and X-rays. In this paper, to find and compare which images will give the better accuracy and prediction of the classification were executed. The separate datasets of Covid and Non-Covid are combined separately for CT and XR images.

### *4.2 Data Preprocessing*

The combined dataset of both Covid and Non-Covid mentioned in the data description is converted into the same format which is 'png' for better classification. Considering the images with different dimensions that are difficult to use and undergone a process for re-sizing results in 128 $\times$ 128 dimensions for all images. This resulted in the dataset is wrapped into a zip file and used for implementation. After unzipping the dataset in the working environment, combine both Covid and Non-Covid images and shuffle them.

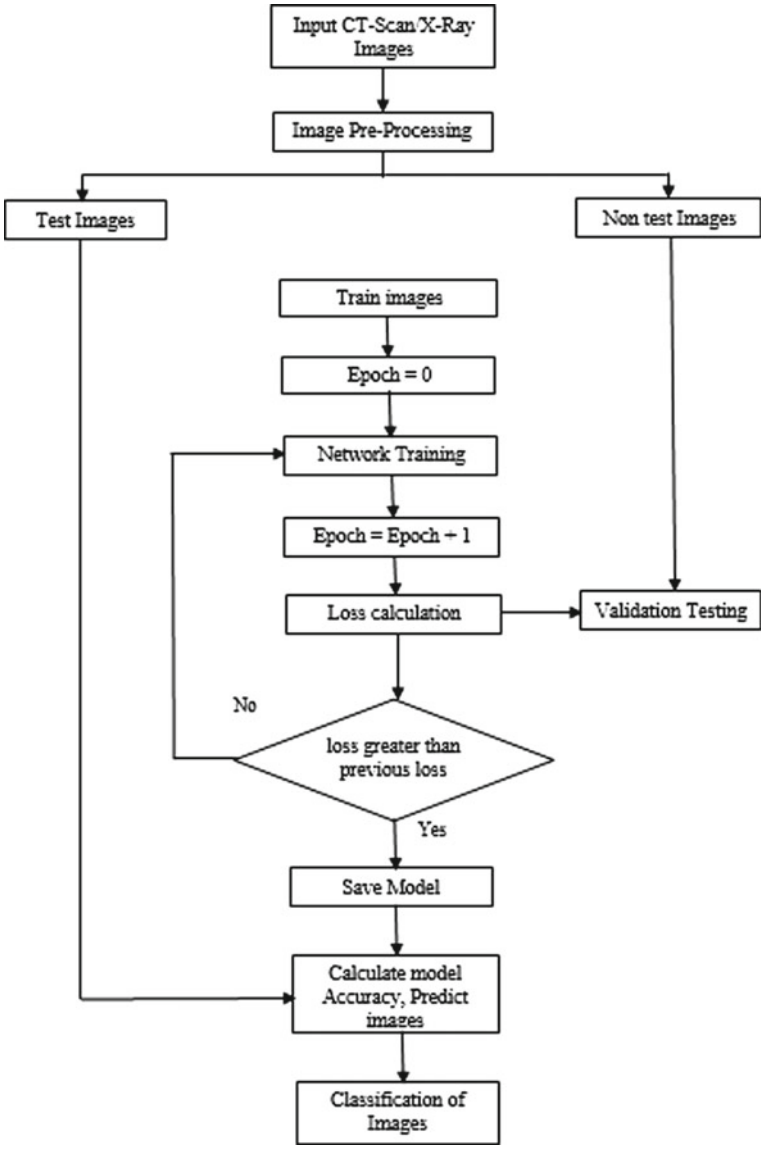



**Fig. 1** Flowchart of working model

## *4.3 Feature Extraction Technique*

In this paper, convolutional neural networks (CNN) are used for feature extraction. The resulted features are sent to the softmax function for classification. From this, it can be determined whether the image belongs to Covid or Non-Covid patients.

### 4.3.1 Convolutional Neural Network

A convolution Neural network is a feature extraction method that contains layers such as a convolutional layer, dense layer, max-pooling layer, and so on. It can have more than one same layer. Various processes take place in layers that extract features of an image. Those features are not visible, but it is possible to classify them based on using other classification techniques (Fig. 2).

## *4.4 Analyzing Results*

Softmax is used to get the score of the predicted model. For evaluating the results obtained from classification, some of the metrics were used in which the commonly used by everyone are precision, recall, $F1$-score, error rate, and accuracy. The confusion matrix is used, which shows the exact number of images classified correctly and incorrectly.

Total is taken as the Number of all the images which are used to classify either correctly or incorrectly classified.

A confusion matrix is represented as cm.

$$\text{cm} = \begin{matrix} x1 & x2 \\ y1 & y2 \end{matrix}$$

$$\text{Total} = \text{cm}[0, 0] + \text{cm}[0, 1] + \text{cm}[1, 0] + \text{cm}[1, 1]$$

Accuracy is calculated by dividing the sum of the correctly classified number of images for Covid and Non-Covid by total.

$$\text{Accuracy} = \frac{\text{cm}[0, 0] + \text{cm}[1, 1]}{\text{Total}}$$

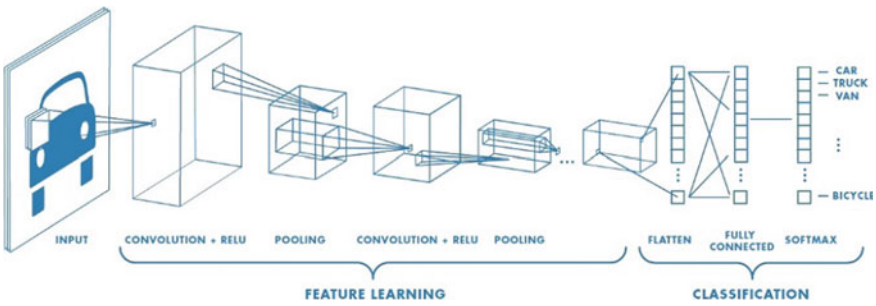



**Fig. 2** Softmax function for classification

Error rate is calculated by dividing the sum of the incorrectly classified number of images as Covid and Non-Covid by total.

$$\text{Error Rate} = \frac{\text{cm}[0, 1] + \text{cm}[1, 0]}{\text{Total}}$$

Precision is calculated by dividing the number of correctly classified Non-Covid by the sum of incorrectly classified Covid images and correctly classified Non-Covid images.

$$\text{Precision} = \frac{\text{cm}[1, 1]}{\text{cm}[0, 1] + \text{cm}[1, 1]}$$

The recall is calculated by dividing the number of correctly classified Covid images by the sum of several correctly classified Covid images and several incorrectly classified as Covid images.

$$\text{Recall} = \frac{\text{cm}[0, 0]}{\text{cm}[0, 0] + \text{cm}[0, 1]}$$

$F1$-Score is taken as the weighted average of recall and precision.

$$F1\text{-Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

## 5 Experimental Results

### *5.1 Results of CT Scan Images*

See Figs. 3 and 4.

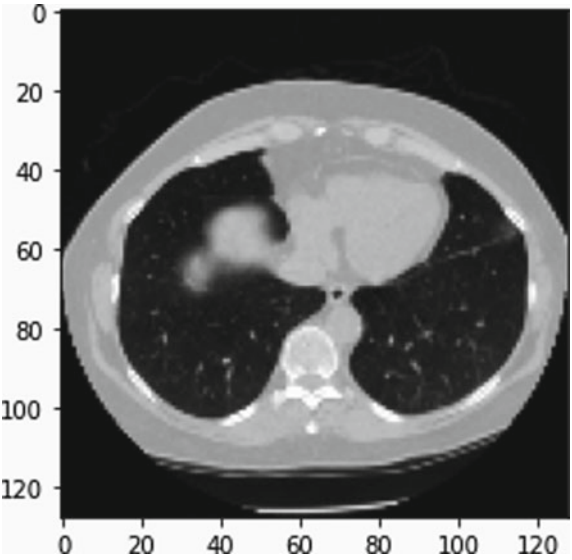
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| COVID | 0.96 | 0.93 | 0.95 | 834 |
| NONCOVID | 0.86 | 0.92 | 0.89 | 375 |
| accuracy | | | 0.93 | 1209 |
| macro avg | 0.91 | 0.93 | 0.92 | 1209 |
| weighted avg | 0.93 | 0.93 | 0.93 | 1209 |

**Fig. 3** Model results

**Fig. 4** Confusion matrix

```
[[777  57]
 [ 30 345]]
acc: 0.9280
sensitivity: 0.9317
specificity: 0.9200
```

**Fig. 5** Non-Covid patient with 73.11%




#### 5.1.1 Result of Input Images

(i) This person most likely is a Non-Covid patient with a 73.11% confidence (Fig. 5).

(ii) This person most likely is a Covid patient with a 66.03% confidence (Fig. 6).

### *5.2 Results of X-Ray Images*

See Figs. 7 and 8.

#### 5.2.1 Result of Input Images

(i) This person most likely is a Covid patient with a 72.26% confidence (Fig. 9).

(ii) This person most likely is a Non-Covid patient with a 53.45% confidence (Fig. 10).

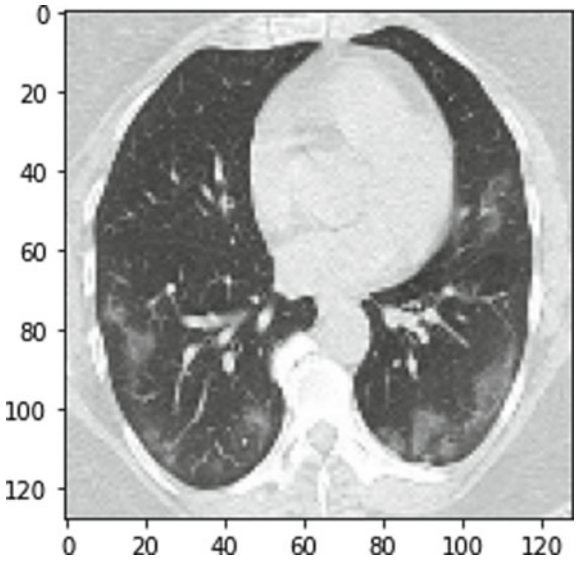



**Fig. 6** Covid patient with 66.03%

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| COVID | 0.82 | 0.74 | 0.78 | 803 |
| NONCOVID | 0.70 | 0.80 | 0.75 | 628 |
| accuracy | | | 0.76 | 1431 |
| macro avg | 0.76 | 0.77 | 0.76 | 1431 |
| weighted avg | 0.77 | 0.76 | 0.76 | 1431 |

**Fig. 7** Model results

**Fig. 8** Confusion matrix

```
[[591 212]
 [126 502]]
acc: 0.7638
sensitivity: 0.7360
specificity: 0.7994
```

The accuracy obtained of 93% using CT scan images and 76% using X-ray images. By comparing the results between CT scan images and X-ray images, it can be concluded that using CT scan images for the classification of Covid-19 is better than using X-ray images.

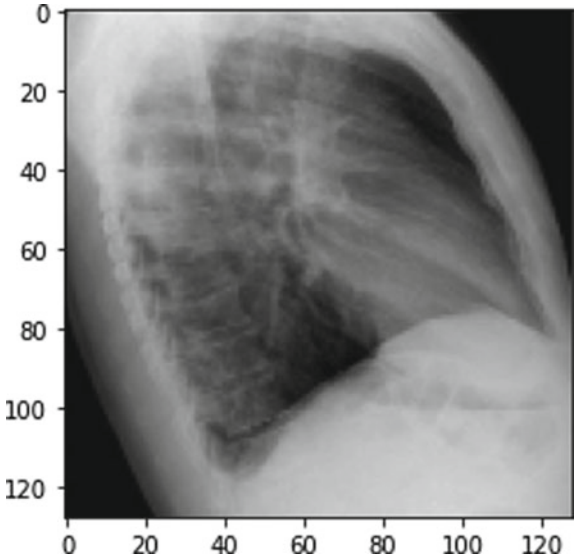



**Fig. 9** Covid patient with 72.26%

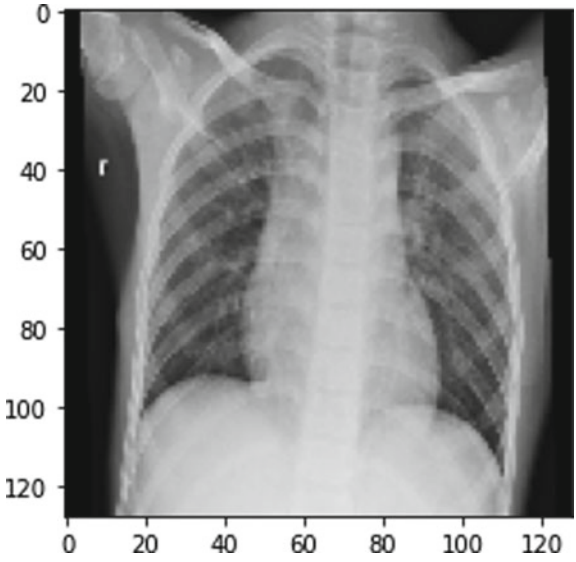



**Fig. 10** Non-Covid patient with 53.45%

## 6 Conclusion

There are many ways to predict Covid-19 from various symptoms such as cough, cold, fever, and tastelessness, and all these aspects are from the past. In this research, an alternative way to predict Covid-19 was suggested. Generally, inputs that are symptoms of Covid-19 were collected in the form of a CSV, i.e., a dataset in which

the findings are not quite reliable. To overcome this difficulty, a new type of dataset consisting of images such as CT scans and X-rays were considered. The amount of precision is much higher than that of the standard CSV dataset. The machine learning and the computer vision concepts like CNN were utilized, and the activation functions like ReLU softmax and CV2 were employed in obtaining high accuracy in the overall Covid-19 prediction. From our findings, it can be concluded that CT scan images are more reliable than X-ray images.

## References

1. Barstugan M, Ozkaya U, Ozturk S Coronavirus (Covid-19) classification using CT images by machine learning methods. https://arxiv.org/abs/2003.09424
2. Öztürk Ş, Özkaya U, Barstu_gan M (2020) Classification of coronavirus (COVID-19) from X-ray and CT images using shrunken features. Int J Imaging Syst Technol:1–11. https://doi.org/10.1002/ima.22469
3. Kassani SH, Kassasni PH, Wesolowski MJ, Schneider KA, Deters R Automatic detection of coronavirus disease (Covid19) in X-Rays and CT images: a machine learning-based approach. https://arxiv.org/abs/2004.10641
4. El-Shafai W, Abd El-Samie F (2020) Extensive COVID-19 X-Ray and CT chest images dataset. Mendeley Data, V3. https://doi.org/10.17632/8h65ywd2jr.3
5. https://datatofish.com/jpeg-to-png-python/
6. https://www.tutorialkart.com/opencv/python/opencv-python-resize-image/
7. Shin H et al (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning. IEEE Trans Med Imaging 35(5):1285–1298. https://doi.org/10.1109/TMI.2016.2528162
8. https://towardsdatascience.com/medical-x-ray-%EF%B8%8F-image-classification-using-convolutional-neural-network-9a6d33b1c2a
9. https://towardsdatascience.com/using-artificial-neural-network-for-image-classification-9df3c34577dd
10. https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8
11. https://www.edge-ai-vision.com/2015/11/using-convolutional-neural-networks-for-image-recognition/

# Johnson's Sequencing for Load Balancing in Multi-Access Edge Computing


**P. Herbert Raj**

**Abstract** Multi-access/mobile edge computing (MEC) is a structural design for enabling cloud computing platform at the edge of mobile network, so as to reduce network congestion, improves fast response, and optimization of mobile resources to compute complex applications. MEC provides a distributed computing environment for its applications. In MEC, all the computational reserves will be brought to the base location of the mobile networks. It has the ability to store and process the contents at physical proximity to the mobile subscribers to assist delay sensitivity and contextaware applications. Generally, the packet forwarding and filtering will be operated by the traditional edge network. At present, applications' online computations and storage are done at remote servers, and those servers are placed far away from the users. New technologies are emerged to shift all the available resources to a edge from a cloud. Radio access network (RAN) could able to do the computational off-loading to the edge of the mobile network. So, it is mandatory to have a load balancer at the edge to distribute all the job to all the available processors without any congestion or delay. A routing specifies how to route the traffic between each origin-destination pair across a network. The traffic sharing is applied in a routing and allocating process to enhance the survivability of a network. The experimental results show the proposed algorithm works better than the existing dynamic load balancing algorithms.

**Keywords** Mobile/multi-access edge computing (MEC) · Flow shop sequencing · Load balancing · Mobile cloud computing (MCC) · Cloud computing (CC) · Make-span and completion time

## 1 Introduction

When IoT devices are connected to the MEC clouds, there will be latency. So, monitoring this latency will be done by latest apps, even though they suffer due to

P. Herbert Raj (✉)
Bandar Seri Begawan, Brunei



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_24

long time travel delay. Because modern apps need to support videos, real-time traffic, and augmented reality (AR). MEC is a pioneering technology that is going to lead the mobile communication technology. Mobile edge computing (MEC) primary job is to move storage after remote storage of cloud to edge.

The mobile users of mid-sized and small-scale industries are gaining profit through mobile cloud computing. It provides a facility to retrieve the data whenever necessary at no matter of time and place. With less or no investment in their storage technology in clouds, these companies can harvest the benefits of protected storage space. Portable devices and users can interact with the clouds, and data management can be done inexpensively in mobile clouds. Ensuring the reliable and authenticated service facility for these portable devices and users is the crucial concern in MCC [1]. The traffic and process will be distributed in order to increase network reliability [2]. The traffic sharing is applied in routing and allocating process to enhance the survivability of network [3]. It is mandatory to devise a routing algorithm to handle all the portable gadgets and mobile users' traffic realistically. So, numerous users' countless requests to access the cloud must be handled properly this algorithm [4].

## 2 Load Balancing

### *2.1 Multi Access/Mobile Computing*

Computation off-loading in MCC for portable devices will be much simpler than expected, but the key issue is latency in partial or full computation off-loading [5]. MEC aims to deliver mobile users a fast response to their request without any delay. Therefore, the local clouds with very few reserves for their processing must handle multiple mobile user requests effectively at a time.

### *2.2 Load Balancing in Multi-Access/Mobile Computing*

This research paper is devised to do load balancing uniformly to all the available resources to improve the usage of the resource and reduce any sort of delay in responding to a user request. A significant job of a load balancing algorithm is to pack an under utilized or over-utilized VMs with number of tasks to get optimum results.

MEC has numerous mobile devices and hosts as depicted in Fig. 1. All the hosts in MEC feed diversified resources to the MEC network's mobile devices, namely RAM and central processing unit for different sort of obligations which includes computation, saving, and optimization. Remotely located clouds are communicated by the hosts of MEC for processing and saving. For example, any environmental data could be fairly collected and transmitted to any one of the hosts in MEC [6]. Host of

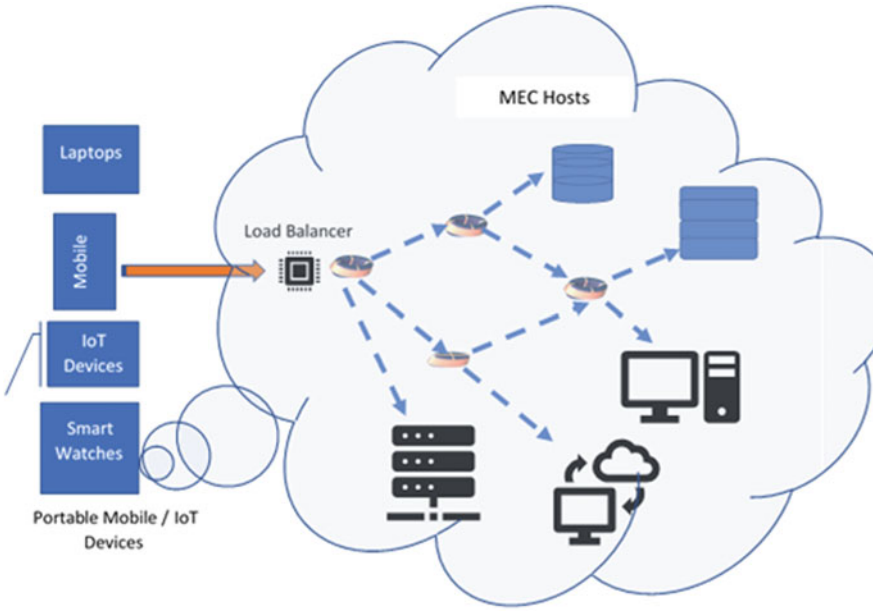



**Fig. 1** MEC model

a MEC collects, examines, and sends the data to a far-flung cloud. MEC hosts which run an analytical video app can able to sense any unusual events after the date stated by cameras, such as burglar, trespasser, lost vehicles to a control center [7].

Resource virtualization is a significant service provided by MEC to other mobile and portable devices. When a mobile device connects to MEC, it does not have any clue about which host of MEC is delivering this service. A mobile edge computing load balancer (MEC-LB) is installed between mobile edge hosts and mobile users. Different sort of apps they need diverse cloud structure namely latency and network bandwidth [8].

## 3 Flow Shop Sequencing

Job sequencing or scheduling is a kind of optimization procedure where jobs or user requests in MEC will be allotted to the requested resources as soon as possible. This will reduce latency and energy consumption [9]. The major goal of this sequencing is to make all the machines busy. There are few conditions that must be considered for this sequencing.

- Job sequence must be orderliness persist equally on each machine.
- The orderliness of job administering on each machine is equivalent.
- Each job must be processed in all the available machines.

### 3.1 Proposed Flow Shop Algorithm for Load Balancing

The given jobs are having even weightage, and user-submitted jobs are arranged in the order in this proposed method. Resources for the underlying networks are provided by the queue manager. Queue manager or load balancer is meant to handle user requests and allocate these requests to available servers. The flow shop load balancing will authenticate the balancing of load and complete resource utilization. The following subdivision 3.2 and 3.3 comprise the approach for load balancing in MEC.

### 3.2 Flow Shop Sequencing for MEC-LB

In job scheduling, resources are allocated with jobs in order to achieve optimization. Consider there are '$n$' jobs or user requests $j_1, j_2, j_3 \ldots j_n$ which are to be allocated to '$m$' machines in MEC hosts $m_1$, $m_2$, $m_3$ … $m_n$. The aim of the job scheduling algorithm is to engage all the machines without being idle. The overall processing spell must be diminished by arranging available jobs properly to all the machines. The aim of this algorithm is to minimize the make-span.

### 3.3 Flow Shop Sequencing Problem

Schedule '$n$' jobs on two machines using the flow shop sequencing technique. The time required by each operation of these jobs is given by the following matrix.

There are totally five jobs, each job should go through Machine $M1$ and Machine $M2$ in the same order with the job handling time as shown in Fig. 2. Here, the total makespan is 52 according to the given job sequence Fig. 3.

**Fig. 2** Job sequence

| Machine / Job | M1 | M2 |
|---|---|---|
| J1 | 10 | 6 |
| J2 | 6 | 12 |
| J3 | 8 | 9 |
| J4 | 8 | 10 |
| J5 | 12 | 5 |

**Fig. 3** Flow shop sequence

| Completion Time | M1 | M2 |
|---|---|---|
| J1 | 10 | 16 |
| J2 | 16 | 28 |
| J3 | 24 | 37 |
| J4 | 32 | 47 |
| J5 | 44 | 52 |
| **Makespan** | | **52** |

| J1 | J2 | J3 | J4 | J5 |
|---|---|---|---|---|

## *3.4 Johnson's Sequencing*

According to Johnson's algorithm, priority will be given to a job with less processing time, and it will be selected; if there is a draw, then select any one of among these jobs. If a job is selected from a first machine, then it will be placed in a first position of the sequence. If a job is selected from a second machine, then job will be placed in a last position of the sequence. The job once allotted in the sequence must be deleted for further sequencing. Repeat the process for other jobs in the list and continue it until all the jobs are sequenced. The processing time of jobs might fluctuate, so the makespan and machine inactive time will be computed by job check-in and check-out policy.

If job sequencing is calculated for the same problem given above according to Johnson's algorithm, it will be resulted in two sequences. These two sequences are shown in the following Figs. 4 and 5. These two sequences yield the same make-span.

**Fig. 4** JA sequence 1

| Completion Time | M1 | M2 |
|---|---|---|
| J2 | 6 | 18 |
| J3 | 14 | 27 |
| J4 | 22 | 37 |
| J1 | 32 | 43 |
| J5 | 44 | 49 |
| **Makespan** | | **49** |

| J2 | J3 | J4 | J1 | J5 |
|---|---|---|---|---|

**Fig. 5** JA sequence 2

| Completion Time | M1 | M2 |
|---|---|---|
| J2 | 6 | 18 |
| J4 | 14 | 28 |
| J3 | 22 | 37 |
| J1 | 32 | 43 |
| J5 | 44 | 49 |
| **Makespan** | | **49** |

| J2 | J4 | J3 | J1 | J5 |
|---|---|---|---|---|

## *3.5 Benefits of the Proposed Methods*

Johnson's algorithm harvests good result than the flow shop sequencing. Johnson's works courteously in two machines, and it can be extended to '$m$' machines ($M > 2$) [10]. Johnson's algorithm not only reduces the make-span but also minimizes the indolent time between servers; moreover, it yields promising optimal results by reducing server response time [11]. Hereby, it concluded that Johnson's algorithm works well for in the MEC-LB. Figure 6 shows the comparison between makespan for flow shop sequencing and Johnson's algorithm.

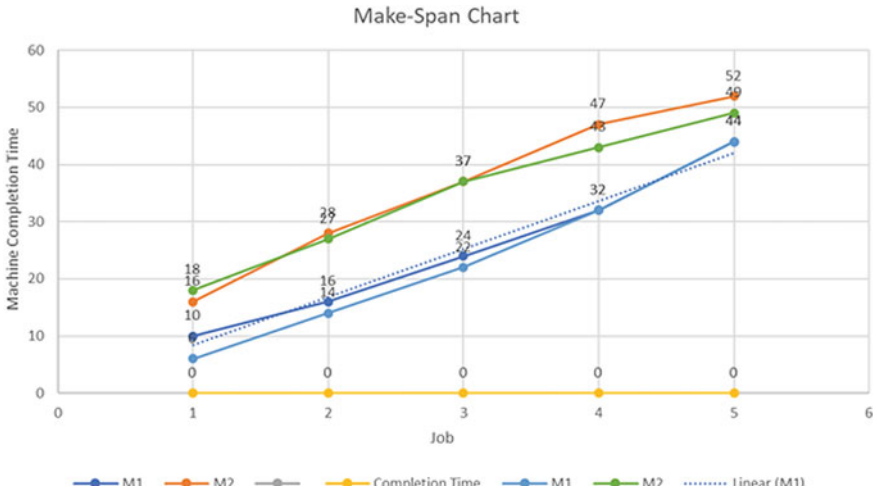



**Fig. 6** Makespan chart

### 3.6 Johnson's Algorithm

- Consider '$n$' number of jobs and '2' machines
- Define, $Ai$ and $Bi$ for $i = 1, 2, 3 \ldots n$ Min $[A_i, B_i]$
- Determine the least time $A_k$ for some $i = k$, process and add the $k$th job at first of a sequence
- Determine the least time $B_r$ for some $i = r$, process and add the $r$th job at last of a sequence
- If there is a tie, $A_k = B_n$, then select any one of the jobs.
- Repeat steps 3 and 4 until all the jobs are assigned in the sequence.

### 3.7 Lemma

Let $X = X(1), \ldots, X(n)$ is a list built by the above given algorithm

| | |
|---|---|
| Then | … $X$ List |
| $U$: $X(1) \ldots X(t)$ | … Left List |
| $R$: $X(t+1) \ldots X(n)$ | … Right List |
| Concatenating $X = UoW := X(1)$ | $\ldots X(n)$ |
| $\min\{Q_{k1}, Q_{m2}\} < \min\{Q_{m1}, Q_{k2}\}$ | … '$k$' job 'm' machine |

denotes that the job k performed earlier than job m in $X$ [12].

### 3.8 Proof

If $Q_{k1} < \min\{Q_{m1}, Q_{k2}\}$, then $Q_{k1} < Q_{k2}$ denotes that the job $k$ goes to $U$.
If $m$ is inserted to $R$, then the job is done.
Else $m$ occurs followed by $k$ in $U$ because of $Q_{k1} < Q_{m1}$.
If $Q_{m2} < \min\{Q_{m1}, s_2\}$ [12].

### 3.9 Advantages

There are numerous advantages of using this type of sequencing in MEC-LB. The significant objective is to minimize the make-span. The primary advantages are minimizing the makespan and augment the utilization of the servers. Achieving nil idle time, maximized machine usage, proficiently progress the process, minimized overall handling time are the superfluous goals of this sequencing method [13].

## 4 Conclusion

In this research article, the proposed Johnson's algorithm considerably reduced the time taken to handle the user requests. This diminishes the network latency. The time complexity of Johnson's algorithm is $O(n \log n)$ [14]. The calculation showed in this article with apt sequencing and routing of machines clearly affirms that flow shop sequencing by using Johnson's rule reduces the total makespan by 5.7%. In MEC-LB, it is a must to optimize the user requests based on the available servers, so that the servers are able to serve the user requests swiftly. The MEC servers' utilization time also will be increased considerably without any idle time. This proposed algorithm accomplishes the significant purpose of MEC by giving information processing amenities at the verge of portable network and users [15]. This research article can be enhanced with calculating the makespan for '$n$' number of jobs and '$n$' number of machines. Few heuristics methods also can be applied here to broaden this research scope.

## References

1. Huerta-Canepa G, Lee D (2010) A virtual cloud computing provider for mobile devices. In: 1st ACM workshop on mobile cloud computing and services: social networks and beyond (MCS), ACM, June 2010
2. Cheng Y, Wang Y, Lu Y, Ji Y, Sun Y (2008) Traffic engineering-based load balancing algorithm in GMPLS networks. Key Laboratory of Optics Communication and Lightwave Technology, Beijing University of Posts and Telecommunications, Beijing, Dec 2008
3. Awduche D, Chiu A, Elwalid A, Widjaja I, Xiao X (2002) Overview and principles of internet traffic engineering, redback networks. Network Working Group, May 2002
4. Sarddar D (2015) A new approach on optimized routing technique for handling multiple requests from multiple devices for mobile cloud computing 3(8):50–61. ISSN: 2321–8363
5. Wei X, Fan J, Lu Z, Ding K (2013) Application scheduling in mobile cloud computing with load balancing. J Appl Math 409539:13. https://doi.org/10.1155/2013/409539
6. Hu YC, Patel M, Sabella D, Sprecher N, Young V (2015) Mobile edge computing a key technology towards 5G. European Telecommunications Standards Institute White Paper
7. Yu Y, Li X, Qian C (2017) SDLB: a scalable and dynamic software load balancer for fog and mobile edge computing, MECOMM '17, 21 Aug 2017. Association for Computing Machinery ISBN 9781-4503-5052-5/17/08, /10.1145 / 3098208. 3098218
8. Herbert Raj P, Ravi Kumar P, Jelciana P (2016) Mobile cloud computing: a survey on challenges and issues. Int J Comput Sci Inf Secur (IJCSIS) 14(12)
9. Srichandana S, Kumar TA, Bibhudatta S (2018) Task scheduling for cloud computing using multi-objective hybrid bacteria foraging algorithm. Future Comput Inform J 3(2):210–230. https://doi.org/10.1016/j.fcij.2018.03.004
10. Joneja A Johnson's algorithm for scheduling. https://ieda.ust.hk/dfaculty/ajay/courses/ieem513/GT/johnson.html
11. Sankar PM, Paramaguru V (2015) Finding an optimal sequence in the flow shop scheduling using Johnson's algorithm. IJISET Int J Innov Sci Eng Technol 2(1)
12. Brucker P Scheduling algorithms, 5th edn. Springer, Berlin, Heidelberg, New York. ISBN 978-3-540-69515-8
13. What is Six Sigma.Net: flow shop sequencing, 2019. https://www.whatissixsigma.net/flow-shop-sequencing/

14. Lee G-C, Hong JM, Choi S-H (2015) Efficient heuristic algorithm for scheduling two-stage hybrid flow shop with sequence-dependent setup times. Math Probl Eng 420308:10. https://doi.org/10.1155/2015/42030
15. Pham Q-V, Fang F, Ha VN, Jalil Piran M, Le M, Le LB, Hwang W-J, Ding Z (2020) A survey of multi-access edge computing in 5G and beyond: fundamentals, technology ıntegration, and state-of-the-art. IEEE Commun Surv Tutor

# A Study on MPLS Vs SD-WAN


**S. Rajagopalan**

**Abstract** Internet service providers and techno enterprises evolved with CISCO IOS called MPLS built a strong and diligent network for succeeding generations to utilize all sorts of full-fledged amnesties over a solitary structure. MPLS is considered as a routing method, and it is not a facility or a service. MPLS can be encased with any prevailing infrastructures, namely digital subscriber line, asynchronous transfer mode, frame relay, and IP. MPLS is not platform dependent. It can work seamlessly without making any change in the current environment of these technologies. But implementing MPLS is quite expensive, so with the support of SD-WAN, ISPs and enterprises are attempted to enhance its usages using inexpensive Internet connection. This survey articled aimed to compare the pros and cons of both the technologies MPLS and SD-WAN.

**Keywords** SD-WAN (software-defined wide area network) · MPLS (multi-protocol label switching) · LSP (label-switched path) · TE (traffic engineering) · ATM (asynchronous transfer mode) · FEC (forwarding equivalence class)

## 1 Introduction

MPLS is a WAN-based system launched to foredetermine the well-established and extremely effective routes to forward data from one node to another node by using short bit sequences called labels instead of lengthy network addresses. A routing method selects a route to direct traffic from source to destination [1]. All the real-time traffic can be handled easily by this method. MPLS technology has been ruling private-based connectivity for the past 20 years. Internet protocols IPv4, IPv6, frame relay, ATM and IPX are all supported by MPLS [2]. SD-WAN is evolved from MPLS technology. SD-WAN guarantees the secured, smooth and secluded connectivity [3]. But MPLS has few issues such as security and backup link. WAN backbone

S. Rajagopalan (✉)
Department of Computational Logistics, Alagappa University, Karaikudi, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_25

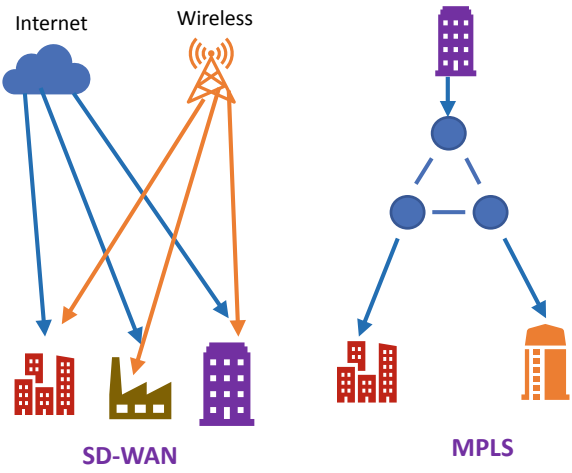



**Fig. 1** SD-WAN and MPLS network. *Source* https://www.fieldengineer.com/sd-wan/sd-wan-vs-mpls-vs-sdn

architecture is completely integrated by the SD-WAN and also drives online traffic by using a centralized strategy. This issue could not be handled properly due to incongruent fragments of its structure and approach. This article discusses those issues in detail in the upcoming sections.

MPLS works alike a router or switch in layer 2.5. Here, data will be forwarded by the decisions of the packet forwarding methods. SD-WAN employed certain WAN connections, namely LTE and 5G which connect large enterprises located in remote places all over the world. Figure 1 illustrates the network connections of MPLS and SD-WAN.

## 2 MPLS

MPLS works independently, and it is not tangled with any of the existing fundamental technologies. The simple MPLS architecture with two sections, namely forwarding (data plane) and control (control plane) is shown in Fig. 2. In MPLS, once a packet arrives, it will be allotted to a forwarding equivalence class (FEC) which is specified by the labels. In the network, every router has a table about neighboring node. So, router will not try to find the header rather the succeeding routers utilize the labels as an index in order to provide new FEC. The job of these path labels is to find out the effective path to destination node instead of finding out endpoints [4]. It was designed to improve the drawbacks in frame relay and ATM [5]. MPLS resides on enterprises and on carrier's infrastructural backbone to support campuses, offices and its branch offices, enterprises and Ethernet-based services in order to achieve QoS for the reliable real-time traffic. The MPLS infrastructure could not fix itself properly with OSI layers. So, it is considered as 2.5 layer [6]. MPLS can establish forwarding tables for any protocol. A label-switched path (LSP) will be established to route

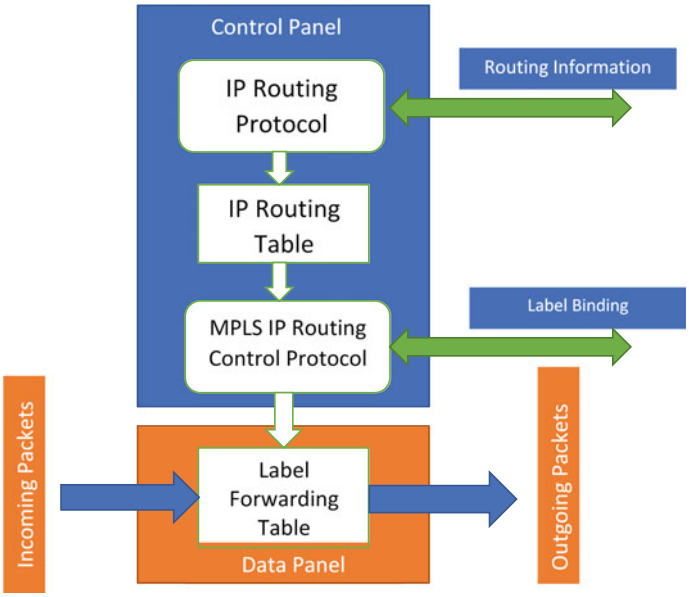



**Fig. 2** MPLS architecture

traffic between foredetermined path in MPLS networks. A unidirectional LSP must be established before the communication starts. Based on this, forwarding would take place. When a user wants to communicate in MPLS network, then MPLS label will be added at the ingress node of the MPLS router. Each MPLS label consists of four parts, namely label, QoS experimental bits, stack bottom and TTL [7]. A label consists of all the details for the MPLS routers to identify the destination of the packet forwarding. Experimental bits are meant for QoS to assure the precedence of the labeled packet. Stack bottom informs MPLS router that the label reaches the egress router, i.e., it reaches the destination.

## *2.1 Boon and Bane*

There are few paybacks in using MPLS, namely bandwidth optimization, congestion reduction, scalability and improved performance. Congestions could be minimized by finding optimal routes [8]. MPLS is not so susceptible to denial of service attack. MPLS is considered to be secured and private transfer mode [9]. MPLS is considered being an expensive mode of transfer because this service must be obtained from a carrier. There is very few MPLS service providers giving global coverage. So, if a company wants to extend its service to its branch office, MPLS communication will be costly and branch office communication to the headquarters will be troublesome

unless they start using clouds. Clouds provide anywhere and anytime services for their users in a cost-effective way [10].

### *2.2 MPLS Future*

ISPs need lot of infrastructural assets to manage the emerging amount of users and applications [11]. MPLS is an essential part of the WAN environment. Nowadays, industries are gradually passing on to a hybrid landscape which has both MPLS and basic Internet service. Point-to-point services still be provided by MPLS for branch offices of big firms, retailers and other data centers. MPLS is very expensive, dependable and consistent but Internet connection is very cheap and reliability is very low. Moreover, incorporation of MPLS-TE with SDN has improved the resource usage and load balancing of a network.

## 3 SD-WAN

Conventional WAN's distinct job is to allow employers in a branch or an institution's campus to link and access the applications installed in their servers. Usually, MPLS's dedicated LSP is used to maintain a secured and authentic connection. This cannot be applied in the present cloud-based scenario. Industries and enterprises started using SaaS and IaaS applications in numerous cloud environments. Present-day IT has lot many contests to be addressed. So IT realizes that the user's application proficiency is complicated and deprived. WAN was intended for the past decades, and it could not handle the current traffic explosion due to the clouds. This caused applications' unreliability, traffic issues and insecured data. Evolution of SD-WAN overpowers these issues.

The era of clouds transmuted backbone of firms. SD-WAN rapidly improves the performance of a business, diminishes budget and provides vigorous security [12]. The SD-WAN architecture is shown in Fig. 3. SD-WAN architecture clearly separated the data plane and control plane and its services. The SD-WAN architecture is cloud scope and protected, flexible and programmable. SD-WAN improves the users' proficiency. It can route IP packets through the most efficient paths, but once it reaches Internet, pursuance is not guaranteed.

### *3.1 Challenges and Benefits*

In the present-day enterprise scenario, over 50% traffic is through clouds but the network available is not at all a cloud serviceable network. SaaS works very pathetically. It has to manage very complicated data flow. The working charges of bandwidth

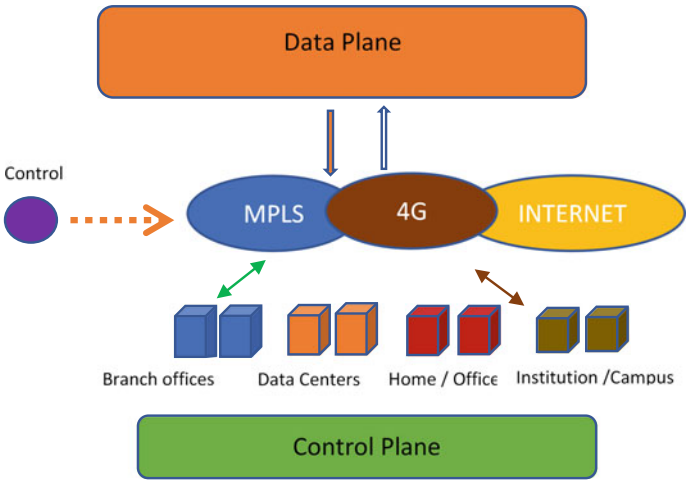



**Fig. 3** Architecture of SD-WAN

will be elevated. Every year, the network traffic flow is increased by 30%. A small architectural change will take months to complete it. User application experience is quite unpredictable, and more than 70% of application disruption is due to network problems. Enterprises' another significant challenge is to implement security. Hybrid architecture's susceptibility leads to many other issues.

Enterprises who would like to migrate to SD-WAN will harvest few benefits. Firstly, they will get optimized cloud and SaaS accessibility and reply times. Any public clouds can be accessed impeccably. Operating cost will be reduced by 50%. Any structural changes will not take long. Any applications SLA will be manageable and easily foreseeable. A rigorous security will be implemented for hybrid networks.

## *3.2 Features*

SD-WAN has lot of features regarding the security, cloud environment, user application experience and cost. User application experience will be very active with hybrid links. In the cloud environment, very vibrant optimization can be done. A data center, cloud and local offices can be connected in single overlay. It can be combined with MPLS and less priced broadband. Due to the base elements of the cloud architecture, it is more vulnerable for any sort of attacks [13]. Cloud security can be provided for hybrid networks when SD-WAN is used. Centralized server system feature is available in SD-WAN. An integrated SD-WAN requires a minimum IT workforce for WAN computerization and segmentation [14].

### 3.3 SD-WAN Load Balancing

Dynamic routing method to find optimum route is automatically done in SD-WAN. When SD-WAN finds the burdened MPLS, it will shift the traffic automatically to the Internet. This makes less congested network route in an inexpensive way [15].

## 4 SD-WAN or MPLS

It is very difficult to determine whether SD-WAN is better than MPLS. Any company routes real-time traffic over WAN, and then the company must need MPLS. SD-WAN utilizes common Internet to connect other Web sites. If an IP packet reaches this common Internet, then there will not be any guarantee of delivery because latency, loss of packet and jitter may affect it. If a company only uses traffic that does not involve real-time applications, namely resource sharing, file sharing, email, file transfer, etc., then SD-WAN would fetch many benefits over MPLS.

### 4.1 High Bandwidth

SD-WAN can assure high-speed Internet through an inexpensive way. It grants a company to use high bandwidth Internet connections. SD-WAN groups several connections to achieve high-speed bandwidth at inexpensive manner.

### 4.2 Better Uptime

Numerous WAN connections can be combined in SD-WAN in order to provide seamless uptime. Cloud services must be always available for the users [16]. MPLS sometimes faces failure to handle this issue.

### 4.3 Increased Performance

SD-WAN regulates the traffic in the fastest way through the circuit. MPLS does not perform any action unless specific settings are made.

**Table 1** SD-WAN and MPLS differences

| | SD-WAN | MPLS |
|---|---|---|
| Types of connections | 5G, LTE and MPLS amalgamation | Twofold MPLS |
| Cloud admittance | Straight and shortest access to clouds | Cloud takes from datacenters |
| Security | Lines are dedicated | Entrenched security |
| Elasticity | Capacity can be increased | Delay in adding the capacity |
| Quality | Business policy based traffic routing | Dedicated lines for Internet traffic |

*Source* https://www.riverbed.com/blogs/sdwan-vs-mpls.html

### *4.4 No More ISP*

SD-WAN does not bump into ISP issue so that it can join or eliminate ISP at anytime without much difficulties. It has the ability to do vibrant link selection [17]. MPLS is struck with same service providers that creates lot of hassle in packet forwarding.

Table 1 focuses on significant differences between the SD-WAN and MPLS.

### *4.5 SD-WAN and Cloud*

If an enterprise uses many cloud-based applications, SD-WAN is the game changer. There are various users who need various services in the cloud environment, and integrating those services is a difficult task in clouds [18]. SD-WAN directs network traffic to a facilitated cloud gateway, and this will connect it to the cloud applications and keep the connection alive [19]. When network flow reaches SD-WAN providers' closest gateway, then it can be straightaway connected to a cloud provider. So the latency, jitter and packet loss will be less comparatively. This will improve the user experience considerably. Secured direct internet access allows branch offices to connect to the cloud applications without any attacks.

## 5 Conclusion

SD-WAN is an upcoming alternative for MPLS, and it makes wide area network connections more adaptable and secured. Any sort of policies can be easily implemented without much changes across the WAN but MPLS always needs a predetermined route. If non-real-time traffic is handled by WAN, then SD-WAN is a perfect choice. If real-time traffic is handled by WAN, then the MPLS must be used. But MPLS connections are quite expensive. Anyway, both kinds of traffic have some

benefits such as high bandwidth, performance improvement, increased uptime, low cost and better performance. SD-WAN can replace MPLS with its centralized trafficking system and gives a more secured and flexible network. Nowadays, nearly 40% of enterprises embraced SD-WAN and 83% of companies attempting to adopt SD-WAN. These numbers may grow in the upcoming years.

## References


1. Kasmir Raja SV, Herbert Raj P (2007) Balanced traffic distribution for MPLS using bin packing method, ICIS, Melbourne University, IEEE, pp 101–106. ISBN: 978-1-4244-1501-4
2. Naganathan ER, Rajagopalan S, Herbert Raj P (2011) Traffic flow analysis model based routing protocol for multi-protocol label switching network. J Comput Sci. ISSN 1549-3636
3. Mehra R, Ghai R, Casemore B (2019) SD-WAN: security, application experience and operational simplicity drivie market growth, Apr 2019. Adapted from Worldwide SD-WAN Survey Special Report, IDC Technology Spotlight
4. Rouse M, Rosencrance L, Sturt R, Scarpati J (2019) Multi-protocol label switching (MPLS), search networking, techtarget, Dec 2019
5. Weinberg N, Johnson JT (2018) What is MPLS: what you need to know about multi-protocol label switching, network world, Mar 2018
6. Rajagopalan S, Naganathan ER, Herbert Raj P (2011) Ant colony optimization based congestion control algorithm for MPLS network, vol 169. In: International conference on high performance architecture and grid computing, HPAGC 2011. Springer Berlin Heidelberg, pp 214–223. ISBN No. Online ISSN 978-3-642-22577-2
7. Lawrence N, Freelance CB (2009) Performance evaluation of MPLS/GMPLS control plane signaling protocols. Blekinge Institute of Technology, Aug 2009
8. Kasmir Raja SV, Herbert Raj P (2007) Identifying congestion hotspots using bayesian networks. Asian J Inf Technol (Medwell J) 6(8):854–858. ISSSN 1682-3915
9. Gary: IPSec and MPLS (even better together), GIAC, SANS Institute (2003)
10. Herbert Raj P, Ravi Kumar P, Jelciana P (2019) Load balancing in mobile cloud computing using bin packing's first fit decreasing method, vol 888. In: Omar S et al (eds) CIIS 2018, AISC. Springer Nature Switzerland AG 2019, pp 97–106. ISBN: 978-3-030-03302-6_9
11. Herbert Raj P, Raja Gopalan S, Padmapriya A, Charles S (2010) Achieving balanced traffic distribution in MPLS networks, vol 8. In: 3rd international conference on computer science and information technology. IEEE, pp 351–355
12. Software-defined WAN (SD-WAN): the new landscape of networking: CISCO (2019)
13. Ravi Kumar P, Herbert Raj P, Jelciana P (2017) Exploring security issues and solutions in cloud computing services. Cybern Inf Technol 17(4):29. Print ISSN: 1311-9702; Online ISSN: 1314-4081
14. Marden BBM (2019) Business value of cisco SD-WAN solutions: studying the results of deployed organizations, CISCO, IDC, Apr 2019
15. Peronkov D (2020) SD-WAN vs MPLS: advantages and challenges, Outlook. https://www.brodynt.com/sd-wan-vs-mpls/, June 2020
16. Herbert Raj P, Ravi Kumar P, Jelciana P (2016) Mobile cloud computing: a survey on challenges and issues. Int J Comput Sci Inf Secur (IJCSIS) 14(12)
17. Riverbed Technologies: What is SD-WAN? 2020
18. Herbert Raj P, Ravi Kumar P, Jelcianan P, Rajagopalan S (2020) Modified first fit decreasing method for mobile clouds. In: 2020 4th international conference on intelligent computing and control systems (ICICCS). IEEE, pp 1107–1110, May 2020
19. CISCO: Optimizing SaaS connectivity using Cisco SD-WAN, Sept 2019

# Security Issues and Solutions in E-Health and Telemedicine


**Deemah AlOsail, Noora Amino, and Nazeeruddin Mohammad**

**Abstract** Telehealth uses wireless communications and Internet services to offer healthcare services remotely. In fact, during pandemics such as COVID-19, telemedicine and e-health services play a crucial role. Security and privacy are of prime importance in e-health systems/services. This paper describes common security issues in telehealth services, such as attacks on end users, medical equipment, and access networks. Further, it also discusses security solutions are employed in telemedicine, including encryption techniques, watermarking, message authentication code (MAC), digital signatures, and the zero trust model. This paper concludes that the use of e-Health services will continue to increase, and security solutions/architectures for these services are of research interest.

**Keywords** E-health · Telemedicine · Security · Privacy · Ransomware · Masquerade attacks · Encryption · Watermarking · Non-repudiation · Zero trust model

## 1 Introduction

Health care is going through numerous transformations. Advancing information and communications technology (ICT) played a major role in the health system's transformation. In the e-health system, the health providers and consumers, policymakers, and researchers can access the healthcare system electronically [1]. Telemedicine is using technology and wireless communication to diagnose, check, treat, and operate on patients via cloud computing [2]. Through telemedicine, patients communicate with medical staff or even medical colleagues to get necessary medical advice. Patients can use telephones or smartphone applications to get their medications. Also, telehealth provides several services to patients. First of all, telemonitoring is to observe the condition of the patient and to monitor the progress or complications that

D. AlOsail (✉) · N. Amino · N. Mohammad
Cybersecurity Center, Department of Computer Engineering, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia
e-mail: 201700348@pmu.edu.sa; 201701727@pmu.edu.sa



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_26

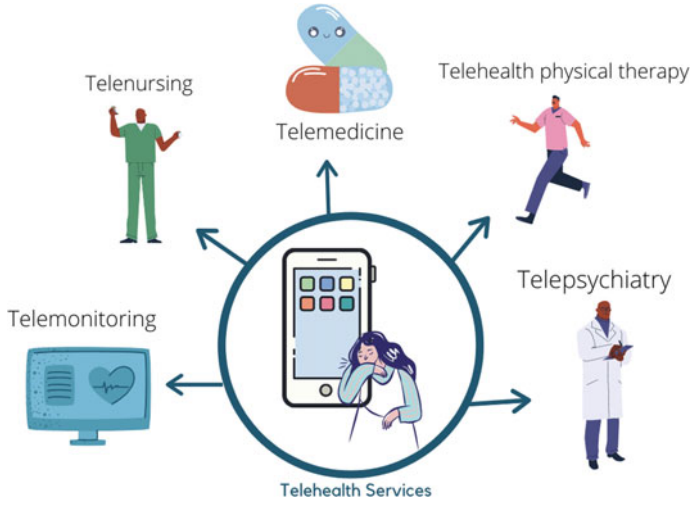



**Fig. 1** Telehealth services

occur to the patient anytime. Secondly, telenursing is where consultations and diagnosing of patients happened. Thirdly, telemedicine involves prescribing the needed medications to the intended patients. Fourthly, physical therapy treatments are where a specialized medical staff offers a physical treatment and makes sure that the patient is following accordingly. Finally, mental health treatment is where sessions are taken online to observe the psychological condition of the patient. Figure 1 lists telehealth services.

The demand for telehealth services has increasingly grown due to the spread of coronavirus disease (COVID-19). Millions of people around the world are infected, and some, unfortunately, faced death. The development of a system to uphold volumes of regular patients as well as COVID-19 patients is necessary to respond to the outrageous disease [3]. With the COVID-19 pandemic, telemedicine's role became significant to manage the challenges that face the health system. Patients are communicating by using applications and Web sites to receive doctors' consultation instead of going to the clinic directly and doctors prescribing medications online. Thus, telemedicine can decrease overcrowding in clinics and emergency units. Furthermore, telemedicine increases health access to those who live in rural areas with minimized care access which can lead to less mortality rates of COVID-19. As reports say, mortality rates due to COVID-19 are higher in areas that have low healthcare access areas in China compared to other areas that have high access [3]. Hence, telemedicine can reduce such burden as Centers for Disease Control and Prevention (CDC), and some other health organizations specified that telemedicine must be present as a healthcare procedure especially during the pandemic which shows the importance of telemedicine [3].

Every week in Saudi Arabia, there are almost 400,000 appointments [4]. Saudi telehealth network's aim is to reach and link remote areas to primary health centers [5]. Thus, it will improve services of the healthcare system, and it will be of high quality [5]. STUE, the Saudi Telemedicine Unit of Excellence, provided some regulations for consultations, such as video consultations, legal requirements same as physical consultations, documenting activities, and training staff [4].

Saudi Ministry of Health provided services for patients, during the COVID-19 pandemic, such as prescription refills online or by telephone calls. Besides, some applications provide telemedicine services for patients to consult doctors by online video calls and doctors prescribe drugs if necessary. For example, Sanar is a Saudi application providing telemedicine services. Security is a crucial factor in the e-health system.

Three crucial fields in security for e-health include availability, confidentiality, and integrity. Moreover, telemedicine includes records and clinical databases, documents' electronic transfer, and data storage and data disposal which are associated with security.

This paper discusses the security issues, considerations, and solutions in telehealth systems. The paper is organized as follows: Sect. 2 describes the major security issues in telehealth systems, masquerade attack, ransomware attack, injection attack, attacks on healthcare cloud system, attacks on implantable and wearable medical devices, home network attack, public network attack, and end user attacks. Section 3 describes security considerations: authentication, integrity, confidentiality, availability, and access control. Section 4 includes security solutions in telehealth, encryption, watermarking, MAC, digital signature, audit, and zero trust model.

## 2 Security Issues

Security issues in health care can cost a huge amount of money. Furthermore, it can cost lives as well, and that is why security problems should be prevented. The main two issues that security targets on e-health and telemedicine are securing the medical data and providing privacy to the patient's information. In e-health, there are several causes for breaches of data that include insider or malicious attacks, lost or stolen devices, and healthcare staff's mistakes. In order to mention the solutions to these issues, we must first address the common cyberattacks on e-health and telemedicine systems. Ransomware and other malware are serious problems in healthcare systems because they may lead to huge consequences such as loss of lives [6].

Hackers tend to target patient healthcare data that includes everything from the medical records, the patient personal details, and even the payment information. One incident happened because a network server hack caused more than 80 million health data records to be stolen. Since the beginning of the year 2014, several major cyberattacks have happened against the healthcare industry and several hundreds

of thousands of patients affected [7]. Clearly, hackers tend to breach the healthcare system in order to get funded but why does healthcare information have a higher demand and cost than credit card information? Some of the reasons are described below.

Healthcare information has a higher value, and it is hard for victims to change or delete their social security number or birthdate in order to protect themselves while the credit card can easily get canceled anytime. Furthermore, the stolen health patient records cost $363 per each and the credit card information often sold for only a few dollars. Moreover, the major cause of these breaches is that organizations do not make use of records' encryption as well as the encryption of data at transit and rest which lead to increasing the number of breaches on the system. As a result, actions beyond the financial aspect could occur, such as purchasing drugs or medical equipment, getting medical care, and other illegal acts [7].

## *2.1 Masquerade Attack*

Hackers from hacktivists and cybercriminals use masquerade attacks to get unauthorized access to the system's personal information and data. They will use a fake identity which could be from utilizing the vulnerabilities on the security system, bypassing the mechanism of authentication, or simply from stolen IDs or passwords.

Moreover, e-health masquerade attacks could occur from internal attackers, like medical staff or from external attackers. After the attacker gets into the e-health system, he\she will have full access to the patient's records and confidential data. Also, the attacker will be able to delete the data and make any modifications even in the network configurations [8].

Furthermore, the masquerade attack can be more dangerous for medical applications based on a wireless sensor network because if a denial of service (DoS) attack was launched, it will cause major and serious disruptions on the healthcare applications in which to make it inaccessible. A masquerade attack could also happen on implantable or wearable medical devices. Figure 2 shows an example hack on the cardiac monitor showing wrong readings of the patient's condition. This can result

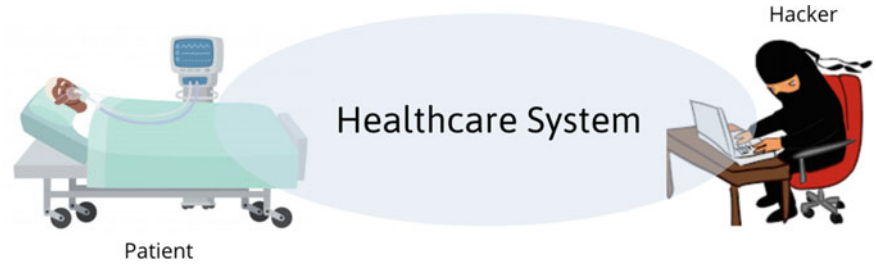



**Fig. 2** Masquerade attack

in performing serious actions by doctors, like giving an overdose or injection that can threaten the patient's life [9].

Additionally, if a masquerade attack gets to capture the patient's physiological data, it can make use of replay threats; where it resends the old treatment messages to the healthcare application again and again causing mistreatment or in other cases overtreatment. These health applications are designed in a way to use the latest medicine messages from the medical sensor networks. Therefore, masquerade attack is considered dangerous to e-health and telemedicine.

## *2.2 Ransomware Attack*

Hackers have found a way to make money by selling healthcare data and records. Therefore, another type of attack arrived called ransomware [8] [DNM1].

Ransomware is a type of malware that prevents the user from accessing the data. It happens by encrypting a key that is known by the hacker, after encrypting the user's data displayed a threat or demand of money, and if the user did not give him the money, the hacker will block the user from accessing data or will publish these data.

On e-health, the hacker will access healthcare providers' records or even the whole hospital records, and this type of attack started in 2016 causing lots of corruption and damage to healthcare data and patient records [8].

## *2.3 Injection Attacks*

One of the major attacks on Web security is the injection attack, where an attacker sends untrusted input to the Web application, and after successful processing it alters the whole program. It is considered as an old dangerous attack where it can lead to data loss or denial of service and in some cases data theft.

SQL injection is a type of the injection attack, which is considered a very dangerous and popular one. It works by injecting an SQL query which after processing targets the SQL database, allowing the attacker to access all databases with the ability to add, delete, or change data [10]. What makes injection attack dangerous is that by only using some tools even inexperienced attackers can try the vulnerabilities of the system and gain control [10].

## *2.4 Attacks on Healthcare Cloud Systems*

E-health systems often use public/private clouds to maintain the medical data of the patients. There are two types of attacks that could occur in a cloud—internal attacks,

which are done by someone who has access to the cloud where the attacker uses his privileges to modify or replace various data within the cloud. On the other hand, an external cloud attack is done by unauthorized users where the attacker uses social engineering and the vulnerabilities in the cloud to attack the transferred data [8].

### *2.5 Attacks on Implantable and Wearable Medical Devices*

Like any other device, medical implantable and wearable devices could be hacked causing several complications to the patient's health, and it could be considered life threatening in some cases. Medical healthcare providers have huge and everyday dependence on electronic devices for diagnosing, monitoring, operating, and communicating in a hospital [11].

For example, an attack could occur on an insulin pump that used to monitor and measure glucose level that could happen only by having some public general information like the radio chip specifications of the insulin pump resulting, eavesdropping on the wireless communications, and after reverse engineering of the protocol and packet format, it will offer full access to the PIN of the remote control. This enables the attacker to have ultimate control, such as showing an incorrect reading, or increasing/decreasing/stopping the insulin injection causing hyperglycemia or hypoglycemia and in some cases death [11].

### *2.6 Home Network Attack*

Telemedicine systems use a patient's network to connect a telemedicine terminal to the central system. Information transmits between the telemedicine terminal at the patient's place (home or office) and the telemedicine system. The patient's network could be Wi-Fi, local area network (LAN), and 4G/5G which will be connected to a telemedicine device, such as blood glucose meter, body fat gage, blood testing, and blood pressure meter [12]. Also, general-purpose operating system (GPOS) embedded devices can communicate by various ways to telemedicine systems. Security threats and attacks, like man in the middle (MITM), could be possible at home network-based telemedicine systems.

### *2.7 Public Network Attack*

The standard Internet services are used to provide communications between a patient's home network and the telemedicine provider. Confidential medical information and prescriptions are transmitted through the publicly accessible network.

Security threats could occur in these kinds of environments by using the vulnerabilities of the telemedicine system which can lead to data sniffing and alterations [12]. Therefore, end-to-end guidelines and encryption of transmitted data must be followed and present.

### *2.8 End User Attacks*

Another factor of security threat is from the end user or more specifically the patient who has no knowledge of the cybersecurity guidelines and principles. Therefore, he/she uses the app not in the proper way and uses simple weak passwords that can attract the hackers on launching security threats, like data theft, phishing, and other software attacks [12].

## 3 Security Considerations in e-Health and Telemedicine Systems

Telemedicine and e-health systems have similar requirements as the general information technology (IT) systems. Some of the requirements are described below.

### *3.1 Authentication*

Authentication refers to truthfulness and originality which means that parties requiring access must be authentic. Authentication issues remain a problem in telemedicine and need solutions. In e-health, for each access, the identities of clients must be confirmed to acquire healthcare data or records [13]. For example, man-in-the-middle attacks could result from incomplete authentication problems in telemedicine. Numerous cryptography algorithms provide endpoint authenticity to especially fight attacks like man-in-the-middle attacks [13].

Cybercriminals including hackers/intruders may also hack the healthcare system to change or delete medical records which can be life threatening [6]. Hackers might steal records to use it to obtain medical services or goods or credit card, and this is called MIDT, medical identity theft [2]. An instance of MIDT was in 2004, and a US electronic medical record was stolen and found in a computer in Malaysia [2]. Thus, authentication of vital information such as medical records or images is a sensitive issue.

### 3.2 Integrity

Integrity is in need in e-Health to provide data consistency and accuracy with the intended information and confirms that data is not altered. By using e-Health, the system must be reliable and without errors. Nonetheless, in the real world, no such thing as an error-free system, so telemedicine, must confront integrity issues such as modification of records [13]. Therefore, integrity issues must be handled and diminished to defeat failures in the system or worse, consequences on patients' health.

As HIPAA states, the importance to implement policies to protect patients' medical records and information by using verification and integrity functionalism [13]. For instance, using hash or checksum on data and if the check on integrity crashes, e-Health application has to report errors and cease the process.

### 3.3 Confidentiality

Confidentiality is critical to security in e-health to keep away unauthorized users from getting private data. Entrusting data in e-health to cloud computing means increased data compromises risks because data is being accessed by a larger number of devices, applications, or parties [13]. The risk of data compromises may lead to patients not entrusting the e-health system and therefore affect the patient–doctor relationship [13]. Furthermore, another risk is that it can deter patients' medical treatment and diagnosis. In addition, data disclosing can lead to dismissing employees who disclosed data. Yet, confidentiality is achievable by encryption and access control [13].

Data remanence is the data's residual presentation that has been erased in a way. As a consequence, data remanence might lead to data confidentiality attacks unintentionally. In telemedicine, when data remanence is not put into consideration, data integrity and confidentiality solutions are inadequate [13].

### 3.4 Availability

Data must be available anytime in the e-health system to avoid critical situations, and this raises the availability issue. Availability includes pursuing medical procedures even when some medical authorities misbehave or during the likelihood of security violations. The system must be available during failures of hardware, denial of service (DoS) attacks, upgrades, and power outages; in other words, the system must have high availability. After administering the HIPAA act on security and privacy, the system has to sustain the usability of e-health data and records [13].

### 3.5 Access Control

It is a method to limit access to patients' public to only entities that are legitimate [13]. Legitimate entities include people given a right to access control authorized by patients themselves or a third party such as medical practitioners [13]. Some solutions are offered for issues related to access control. The most common models for e-health cloud applications are ABAC and RBAC, attribute-based access control, and role-based access control, respectively [13].

## 4 Solutions

There are several healthcare practices that could be used to reduce the possibility of potential risks. Accordingly, most solutions deal with a part of the problem instead of completely solving the issues at once. Derriford Hospital in Plymouth, plastic surgery department, used the store-and-forward telemedicine method and found that concerns relating to e-health including confidentiality, security, and risk were solvable and suggested developing more telemedicine services [14]. Kvedar et al., recommend improving and investing to develop telemedicine to increase the quality of the healthcare system. HIPAA, Health Insurance Portability and Accountability Act, in 1996 detects and corrects errors related to security in telemedicine and improves the healthcare system [1].

Security services including firewall, antivirus, and VPN are implemented in healthcare organizations. In Fig. 3, a study stated 81% of healthcare organizations

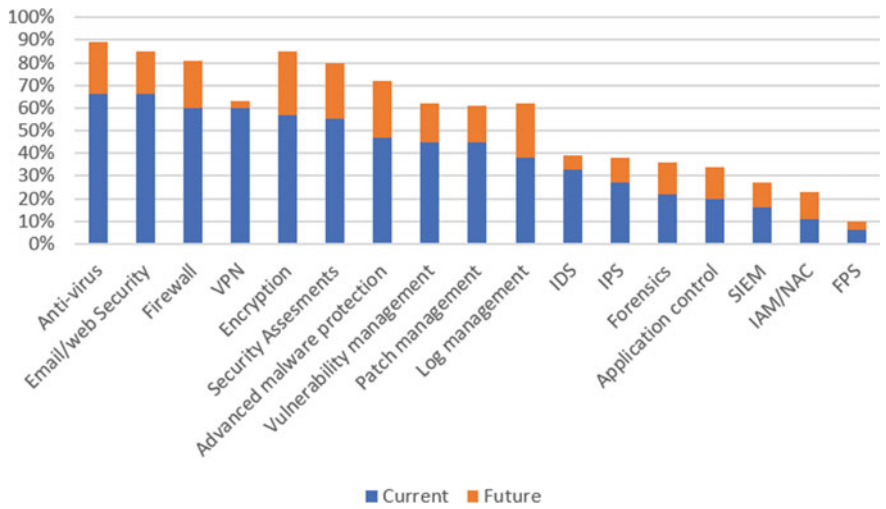



**Fig. 3** Security services of healthcare organizations [15]

are either using firewalls or plan to use it in the future. These security services are basic, and more high security services need to be implemented such as SIEM and NAC. As shown in Fig. 3, some healthcare organizations started to implement SIEM and NAC security services.

In the following subsections, some key security mechanisms are described.

## *4.1 Encryption*

An old and effective mechanism to protect data is encryption. When used properly, it protects the healthcare data at rest and in transmission. Al-Issa et al. suggested data encryption and showed that data encrypted as ciphertext is more secure compared to non-encrypted data and prevents unauthorized users to some extent [13]. Furthermore, encryption can provide secure transmission, even though the attacker can analyze the size and timing in network packets by using traffic analysis, which is called side-channel attacks. Accordingly, adding delays may solve the problem of side-channel -attacks.

Transport layer security (TLS) is used to increase communication security in Web applications by reserving an encrypted medium between a sender and a receiver that sends the encrypted text and transfers the key used for encryption via public key cryptographic protocol. Healthcare data that is kept on the public/private cloud can suffer from data-leaking risks [13]. Consequently, self-encrypting drive (SED) can offer automatic encryption and decryption of data in rest. It is highly significant to control and store encryption keys securely in external means. Additionally, data deletion encryptions secure deleted data and make it unrecovered by attackers.

Nevertheless, when data is decrypted by the receiver, the data protection ends, because it can be copied and forwarded using other means by the receiver. Thus, encryption is not sufficient in protecting sensitive data that risks patients' confidentiality [2]. Figure 4 shows how decrypting a CT scan leads to loss of protection and thus might be copied and distributed to others.

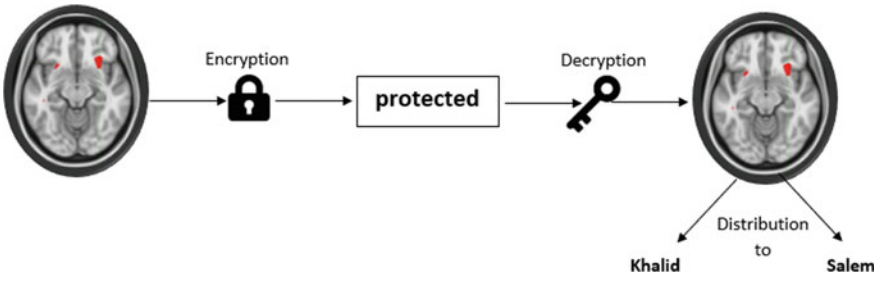



**Fig. 4** CT scan unprotected after decryption [2]

### *4.2 Watermarking*

Watermarking is a solution to decrease authentication issues in telemedicine. Digital watermarking is used to fight against piracy, and this raised authentication and integrity of media. Watermarking must guarantee not interfering with the diagnosis in digital images such as mammograms. Additionally, watermarking tracks illegal copies of media and prevents changes in digital media. In fact, watermarking is preferred over other methods that protect data authentication and integrity and that is because of its critical qualities. Some of the watermarking's qualities include the fact that the watermark is imperceptible and inseparable of the content which improves the security of the data [2].

Furthermore, digital watermarking is used to authenticate images, protect copyright, detect corruption, and monitor data. The host signal is the object communicating, and digital watermarking delivers more information about it. Therefore, digital watermarking is used in numerous applications, and it can conceal plaintext, images, information encoded, or identifications [2].

The system of digital watermarking has two main blocks: The first one is the embedding block which puts the information of the watermark. On the other hand, the extractor block extracts the information that has been watermarked. Besides, robust watermarks are required in the system as an algorithm to resist attacks. Methods of watermarking, that detects damage of corrupted images, authenticate images to increase security, integrity, and confidentiality, are not widely used in telemedicine [2]. As mentioned above, watermarking can increase the security and privacy of data. Hence, organizations developing telemedicine should consider digital watermarking.

### *4.3 Message Authentication Code (MAC)*

Message authentication code (MAC) is a block that consists of some bytes that are added on the message, and then the receiver can check it and verify the authenticity of the data. So, MAC is based on sharing a key between the sender and receiver. The main purpose of MAC is to ensure the authenticity and integrity of the message and that there is no one modified it. Moreover, in case any modifications happen, figuring that will be easy [10]. When A and B are communicating, an eavesdropper can listen to the message and MAC; however, he/she will not be able to change the integrity of the message because the eavesdropper does not have the key. Also, MAC has a property that states that MAC cannot be repeated on any other message. Therefore, it can be a solution to masquerade attacks or any integrity issues in telemedicine and e-health services.

### 4.4 Digital Signature

Non-repudiation is the act of not denying things. In other words, ensuring that a user cannot deny the authenticity of their signature or in a message they already sent. This act is useful in E-communication in order to ensure message transactions between participants [16]. So, if a user signs a contract, he cannot refuse the terms that were on that contract, and he needs to take the responsibility accordingly. Digital signatures can be an example of non-repudiation since it ensures the authenticity of the user and integrity of documents (e.g., telemedicine prescriptions and medical reports) that they have not been changed or modified by any party.

### 4.5 Audit

Auditing can help secure the e-health system, and it logs activities of users in the system chronologically as in recording each access and changes to data. HIPAA act holds health providers responsible for handling protected information of patients [13]. Auditing approaches may lessen insider threats by detecting unauthorized users and illegal leakage of medical information. Likewise, auditing can support detecting hackers' attempts to interrupt the telemedicine cloud system and help to find weaknesses in the system [13].

### 4.6 Zero Trust Model

Zero trust model was developed in 2004, and as its name indicates, it has no trust, so it always verifies [17]. Thus, the zero trust model assumes there is a breach so it tends to verify each request. Instead of protecting network segments, the zero trust model protects resources. The strategy of the zero trust model is to show authenticated users what and only what they are allowed to access. The principle of zero trust model is to treat it as the Internet, assuming there are regular and malicious users on the network. The network, application, and database do not trust any action except when it is identified, approved, and authenticated [17]. Nonetheless, this model works well when organizations know their users and their interaction with other devices.

To use the zero trust model in telehealth, the management of assets and devices such as IoT devices, computers, and surgical robots is vital. However, managing all assets manually is not possible, when the network size is big. Therefore, a deployed authentication service that identifies devices or users on connection by connection basis to control access is imperative [17]. The zero trust model in telehealth must focus on device identity and health and access control. For instance, when hackers access the network with stolen credentials, the attack will not proliferate through the network. Thus, the zero trust model in telehealth can decrease the impact of attackers

by reducing the opportunity of malicious users to explore and breach interactions between assets and users because requests and identities are monitored continuously and attackers need to have privileges in every step that makes it more difficult. Furthermore, zero trust makes it easier for the healthcare environment to detect and solve small failures and breaches because often major attacks begin with minor endpoint exploits. Additionally, in the zero trust model, every step done by the attacker is logged and can be investigated later. Such extensive logging makes attackers vulnerable to detection [17]. Although telehealth has a long way to adjust and shift to a system that uses a zero trust model, it should implement a zero trust model because it reduces serious risks. In telehealth, building a strong basis using a zero trust model starts with awareness of people, identity, authentication, and groups. Over time, the zero trust model will save resources and money by preventing the attacks. For example, the US healthcare system spent around $160 million to recover from ransomware as reported in [17].

## 5 Conclusion

Telemedicine and telehealth are considered a useful way of communication between patients and physicians, especially if both were not at the same place and time. It can increase the quality and access to health care, reduces cost, and provides patient engagement and satisfaction. Furthermore, using electronics in health care can also improve communication, patient care, and safety. However, security and privacy issues can develop in such interactions. Research work discussed some of the cybersecurity issues that could occur in e-health and telemedicine systems. Several attacks, such as masquerade, ransomware, replay, and injection, could occur causing significant damage and corruption to both healthcare data records and patient's personal sensitive information. The research also mentioned some of the proper solutions to these problems, like watermarking, message authentication code, digital signatures, audit, and zero trust model by which the advantages of the new technology and reduced potential risks from happening could be obtained in the future. A code of practice needs to be implemented and followed by healthcare staff to protect and maintain a secure environment.

## References

1. Kvedar J, Coye MJ, Everett W (2014) Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff 33(2):194–199. https://doi.org/10.1377/hlthaff.2013.0992
2. Olanrewaju RF, Ali NB, Khalifa O, Manaf A (2013) ICT in telemedicine: conquering privacy and security. Retrieved 18 July 2020 from https://ejcsit.uniten.edu.my/index.php/ejcsit/article/view/39/28

3. Rockwell KL, Gilroy AS (2020) Incorporating telemedicine as part of COVID-19 outbreak response systems. Am J Manag Care 26(4):147–148. https://doi.org/10.37765/ajmc.2020.42784
4. Telemedicine, an opportunity to maintain continuity of care for outpatients (2020, May 18). Retrieved 19 Aug 2020 from https://home.kpmg/sa/en/home/insights/2020/05/telemedicine-in-saudi-arabia-an-opportunity-to-maintain-continuity-of-care-for-outpatients.html
5. Saudi Telehealth Network. (n.d.). Retrieved 19 Aug 2020 from https://nhic.gov.sa/en/Initiatives/Pages/communicationmedicine.aspx
6. Top 10 threats to healthcare security (2018, Jan 08). Retrieved 13 July 2020 from https://resources.infosecinstitute.com/top-10-threats-healthcare-security/
7. Hackers selling healthcare data in the black market (2015, Aug 10). Retrieved from https://resources.infosecinstitute.com/hackers-selling-healthcare-data-in-the-black-market/
8. Zeadally S, Isaac JT, Baig Z (2016) Security attacks and solutions in electronic health (e-health) systems. J Med Syst 40(12):263
9. Can your cardiac device be hacked? (2018, Feb 20). Retrieved from https://www.acc.org/about-acc/press-releases/2018/02/20/13/57/can-your-cardiac-device-be-hacked
10. Liu D (2011) Next generation SSH2 implementation: securing data in motion. Syngress
11. Li C, Raghunathan A, Jha NK (2011, June) Hijacking an insulin pump: security attacks and defenses for a diabetes therapy system. In: 2011 IEEE 13th international conference on e-health networking, applications and services. IEEE, pp 150–156
12. Camara C, Peris-Lopez P, Tapiador JE (2015) Security and privacy issues in implantable medical devices: a comprehensive survey. J Biomed Inform 55:272–289
13. Al-Issa Y, Ottom MA, Tamrawi A (2019) EHealth cloud security challenges: a survey. J Healthcare Eng 2019:1–15. https://doi.org/10.1155/2019/7516035
14. Wallace S, Sibson L, Stanberry B, Waters D, Goodall P, Jones R, Evans J, Dunn R (1999) The legal and risk management conundrum of telemedicine. J Telemed Telecare 5(1_suppl):8–9. https://doi.org/10.1258/1357633991932748
15. Snell E (2016, Dec 14) How Ransomware affects hospital data security. Retrieved 25 Aug 2020 from https://healthitsecurity.com/features/how-ransomware-affects-hospital-data-security
16. Nurhaida I, Ramayanti D, Riesaputra R (2017) Digital signature and encryption implementation for increasing authentication, integrity, security and data non-repudiation 4:4–14
17. Davis J (2020, Aug 04) How zero trust in healthcare can keep pace with the threat landscape. Retrieved 25 Aug 2020 from https://healthitsecurity.com/features/how-zero-trust-in-healthcare-can-keep-pace-with-the-threat-landscape

# Accident Alert System with False Alarm Switch


**S. Alen, U. Advaith, Joveal K. Johnson, Kesia Mary Joies, Rahul Sunil, Aswathy Ravikumar, and Jisha John**

**Abstract** In this fast-paced world, people are busy chasing their lives. The stress, burden, and carelessness have caused different accidents on roads. Over speeding is one of the reasons for accidents which have caused a huge loss of lives due to the lack of immediate treatment. This research work provides an optimum solution to this problem. According to the proposed work, turbulence on the vehicle will be detected by the accelerometer and can be deactivated within the initial 20 s period using false switch alarm in case it is a false emergency. The accident location and other details of the driver will be sent to the rescue team. The paper proposes a system where immediate alert can be given to authorities and thus save a person injured in the accident within minimum time.

**Keywords** Accelerometer · GSM module · GPS module · False alarm switch · Accident detection · Alerting system

## 1 Introduction

According to the latest reports, India accounts for about 5 lakh road accidents annually, one of the highest in the world, in which about 1.5 lakh people die and another 3-lakh become crippled [1]. Despite the number of awareness sessions organized by different organizations (governmental as well as non-governmental) about the speed limits and other measures against careless driving, accidents are taking place now and then. However, many lives could have been saved if the accident was alerted

S. Alen · U. Advaith · J. K. Johnson · K. M. Joies · R. Sunil · A. Ravikumar (✉) · J. John
Mar Baselios College of Engineering and Technology, Mar Ivanios Vidya Nagar, Nalanchira, Thiruvananthapuram, Kerala 695015, India
e-mail: aswathy.ravikumar2019@vitstudent.ac.in

R. Sunil
e-mail: rahul.sunil@mbcet.ac.in

J. John
e-mail: jisha.john@mbcet.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_27

to the emergency service on time. Conventionally, an accident is informed to the police or hospital only when a concerned citizen passes along that route and takes an effort to inform the authorities. Therefore, an efficient automatic accident detection system that notifies emergency service immediately with the accident location which is a prime need to save lives. Most of the accident alert systems are unable to detect the seriousness of the accident, thereby sending false alarms to emergency service. This has caused authorities to ignore even serious accidents. An accident is an unpredictable and unintentional event. Prompt assistance to those injured in accidents could make a significant difference in the consequences of such accidents and has a high possibility of saving one's life. This system helps in the early detection of accidents by detecting the turbulence on the vehicle and communicating the information immediately to the emergency responses on time to provide quick assistance for the injured person.

## 2 Existing Works

Accident detection and alert systems have been a major study in the past few years. Many research works are going now and then to reduce the number of road accidents. The existing system, which is proposed to reduce accidents, consist of an android smartphone and an accident detection system which checks whether the vehicle is in normal posture or in fallen conditions then it checks the heartbeat of the person and if finds any abnormalities sends a message through an proposed android application via Bluetooth and alerts the rescue team [2]. Many accident cases finally end up in death [3], due to proper medical assistance at the right time. A GPS and GSM module [4] is used to track the accident spot using Google Map service, as an accident occurs, a crash is detected and alerts the nearby police station and the medical team so that the rescuing operation would be more effective [4]. In other systems, messages are sent using the ZIGBEE module and location is sent through a GPS module. Accident is detected using a microelectromechanism (MEMS) [5] sensor and vibrator sensor [6]. ZIGBEE is used to send the message to the ambulance and GPS module to identify the location of accidents [3] and alerts the rescue team for Speedy recovery. The limitations to this existing work is that, it is not cost-effective. An accelerometer is used in detecting accidents in some proposed works [7]. When an accident occurs, there would be a sudden change in the acceleration of the vehicle and an accident is detected using a microcontroller and sends a message to the emergency [8] medical team via GSM module [9]. All the other existing works did not refer to any kind of false alarm system which can avoid creating a panic situation. We should try to avoid such false alarms, so in the proposed system we have added a false alarm switch to send an alert if the person is safe [6]. This would help to ensure the proper use of the resources and manpower of the medical team. Sometimes these false alarms may panic the closed ones to the person involved with the help of this false alarm switch, we can avoid such situations.

## 3 Proposed System

The proposed system deals with the detection and alerting of accidents, by detecting the range of impact upon the car. When turbulence or crash occurs to the vehicle, there would be huge variations in the data provided by the accelerometer. The data of the accelerometer is sent to the microcontroller and if the values are above threshold values (confirming accident), an alerting message is sent to the concerned authorities and some personal contacts within 20 s unless it is canceled by a false alarm switch. False accident alerts can hence be overcome by this proposal. The system design of accident detection and alert system is based on various Arduino Mega modules (Fig. 1).

### *3.1 Technical Specification*

See Table 1.

### *3.2 Implementation*

See Table 2.

When turbulence or any accident is met by the vehicle, it is detected by an accelerometer. The accelerometer is used to detect the sudden change in any value in the axis. With signals from an accelerometer, a severe accident can be recognized. When the value exceeds the programmed maximum limit or threshold value, the alert system gets activated. Once the alert system is activated, it triggers a red LED and the buzzer starts ringing. The 20 s timer also starts at this point. After this, two processes happen asynchronously. The system acquires the current location of the

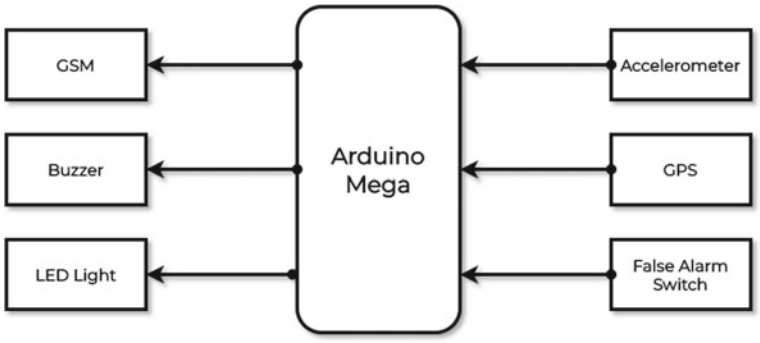



**Fig. 1** Proposed architecture

**Table 1** Proposed architecture technical specifications

| S. No. | Hardware components | Technical specification |
|---|---|---|
| 1 | Arduino mega | Operating voltage 5V<br>Input voltage 7-12V<br>Digital I/O pins 54<br>Analog input pins 16<br>Clock speed 16 MHz<br>Length 101.52 mm<br>Width 53.3 mm<br>Weight 37 g |
| 2 | GPS module Neo6M | Navigation update rate 5Hz<br>Default baud rate 9600bps<br>Sensitivity -160dBm<br>Supply voltage 3.6V<br>Maximum DC current 10mA<br>**Operation limits**<br>Gravity 4g<br>Altitude 50000m<br>Velocity 500m/s |
| 3 | GSM SIM900 module | Dual-Band 900/ 1800 MHz<br>Dimensions 24x24x3 mm<br>Weight 3.4g<br>Voltage range 3.2 |
| 4 | Accelerometer ADXL335 | Operating Voltage 1.8V–3.6 V<br>Operating Current 350µA<br>Sensing Range ±3g<br>Sensing axis 3 axis<br>Sensitivity 330mV/g<br>Shock Resistance 10,000g<br>Dimension 4x4x1.45mm |

device and sends it to the primary emergency contact which are your closed ones, so that they can check up on you. It waits for the 20 s timeframe to check if the false alarm switch is initiated and confirmed, if not, it will send a message along with the location to emergency contacts, nearby hospital as well as a police station (Figs. 2 and 3).

**Table 2** Proposed architecture component uses

| S. No. | Hardware components | Description |
|---|---|---|
| 1 | Arduino mega | For interfacing all the components. It receives the sensor values from the accelerometer where it is evaluated to check whether the accident alert system has to be triggered |
| 2 | GPS module Neo6M | To acquire the latitude and longitude of the accident location |
| 3 | GSM SIM900 module | For sending out the emergency message to the contacts along with the location data |
| 4 | Accelerometer ADXL335 | For measuring acceleration, velocity, orientation, displacement of the vehicle which gives analog data as output |
| 5 | False alarm switch (FAS) | To disable the system by the driver incase of any false alarm |
| 6 | Buzzer | To alert the driver if the system detects any sudden turbulence by producing a beeping sound |
| 7 | LED | To alert the driver of any sudden turbulence on the system by emitting a red light |

## *3.3 False Alarm Switch*

When the alert system gets activated, it initially triggers the led to start blinking and the buzzer to start ringing. After which, it will start the 20 s waiting period, during which a message along with the current location is initially sent to the rider's primary emergency contact so that he can check up on the rider even if it is a false alarm or not. If the system got activated by mistake, there is an option for the rider to cancel the system from alerting the hospital or police station. In order to do this, there is a false alarm switch (FAS) which the rider can switch on to activate it. Once it is switched on, a timer for 5 s will start to confirm the decision. At the end of 5 s, the buzzer will stop ringing and the rider is supposed to switch off to confirm his/her decision. The false switch is designed in this way so that it does not get canceled due to the impact of the accident. If the FAS is not switched off within the initial 20 s period, then it will not be considered and the message will be sent to emergency contacts, nearby hospitals and police stations. If the FAS is turned off within the initial 20 s period, the LED light will also stop confirming the rider's decision to deactivate the alerting system. The reason why we chose 20 s was based on a study [8, 10, 11] which emphasized on the importance of low response time which is very crucial in car accidents (Fig. 4).

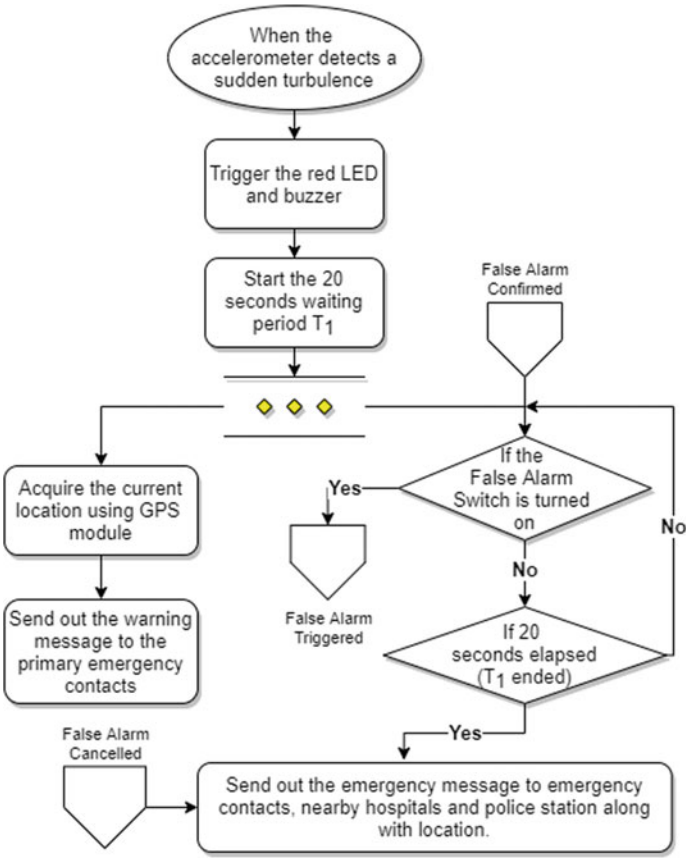



**Fig. 2** Workflow of proposed work

## *3.4 Arduino Mega*

During the prototyping stage, we switched to Arduino Mega from Raspberry Pi due to several reasons such as: Arduino is more energy-efficient than Raspberry Pi. An Arduino can run even on the car's charging port which keeps the voltage above a certain level, along with a primary shield to manage the power but for running a Raspberry Pi, and it requires an unfluctuating supply of 5 V. Even if the power drops on the Arduino, it won't end up with a corrupt operating system or any software errors but in the case of Raspberry Pi, if it is not shut down within the operating system like any other computer, there can be corruption and software problems. The absence of analog pins in Raspberry Pi The accelerometer sensor in our system outputs analog data, so our central board should be able to work with analog data seamlessly. Raspberry Pi is a digital-only computer, so it does not have any analog inputs. In order for the Raspberry Pi to read an analog input, we need to use an

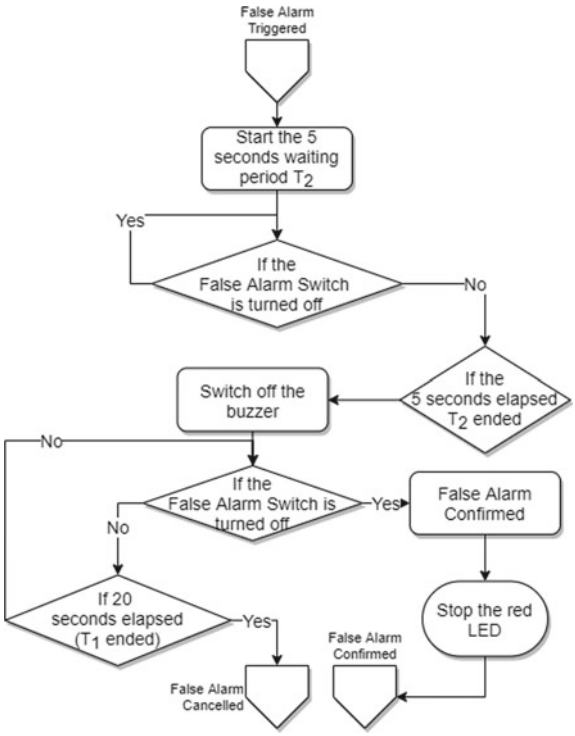



**Fig. 3** Workflow of false alarm switch

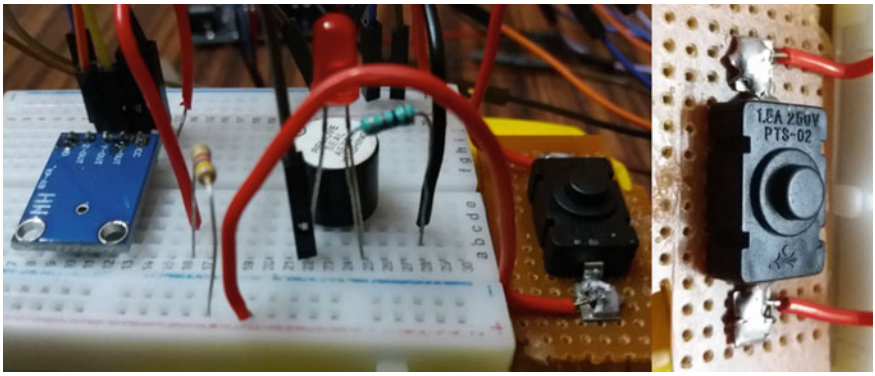



**Fig. 4** False alarm switch

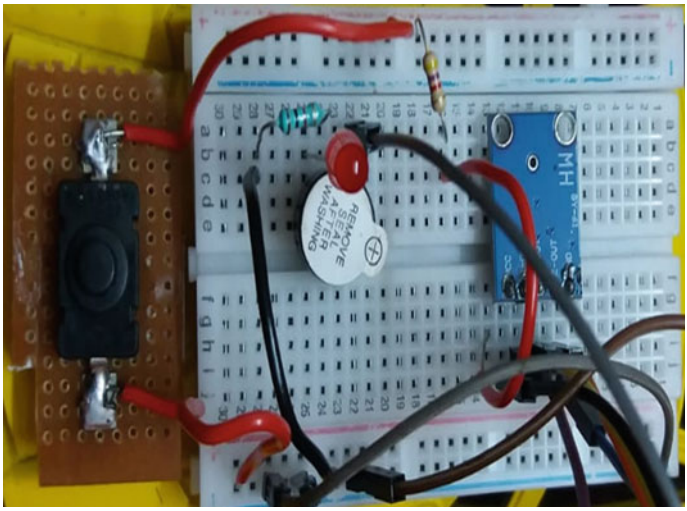



**Fig. 5** Proposed system sensor unit + false alarm switch

analog-to-digital converter (ADC) which adds unwanted complexity to the circuit design along with making it bulkier. While on the other hand, the Arduino Mega comes with 16 analog inputs which can be used directly to read the analog input from the sensors. Simpler design since our system is primarily hardware-based, it does not need all the complexity of running an OS for the system to keep running. This will also help us in the future during the production stage, during which we'll be assembling the system by ourselves [12].

## 4 Result

The system was tested on a very controlled environment, and the results produced are presented here (Figs. 5 and 6).

## 5 Conclusion and Future Scope

The proposed accident alert system plays a significant role in the field of accidents. This device will be really useful in saving the lives of many who are injured in accidents by alerting the concerned ones in time. If an accident occurs in an isolated area, the device would be really helpful in reporting it and hence improve the chances of faster medical treatment to the victim. This system detects accidents based on impact, and it can be extended by adding computer vision to identify the accident from

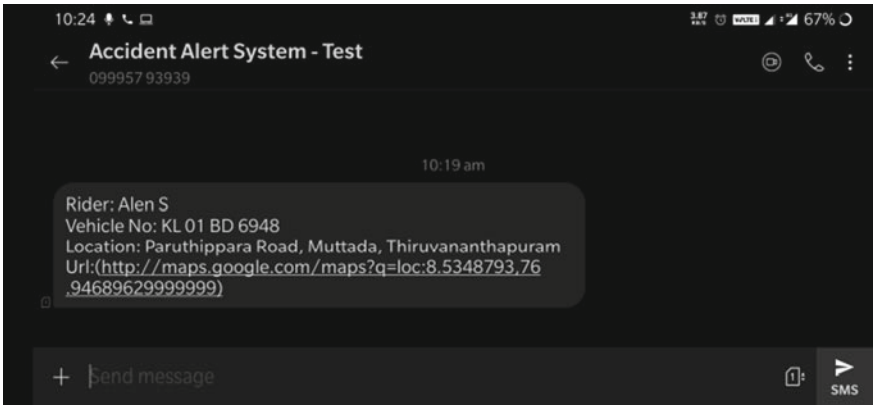



**Fig. 6** Accident alert system

the images of the surroundings captured by a camera. A combination of accelerometer and vibration sensor gives more accurate results of the impact. The compact version of this system could be incorporated into different types of vehicles as well as in helmets. It can be commercially distributed to vehicle and helmet manufacturers. A mobile application will be built integrated with this system, giving a detailed analysis of accidents rate, along with precautions and driving rules. The driving pattern of the user will be studied and can be used for further purposes. The project will be beneficial to the automobile industry and will help society. In the proposed solution, the emergency contacts or information of the authorities needs to be stored in the user's device, which arises an issue as the system may not be useful in certain cities. The system can be revamped by sending the accident location directly to the server, from which the details can be shared with the closest hospitals or emergency services located in that city.

## References

1. Overview of road accidents in India by PRS legislative research. https://www.prsindia.org/policy/vital-stats/overview-road-accidents-india
2. Kattukkaran N, George A, Haridas TPM (2017). Intelligent accident detection and alert system for emergency medical assistance. In: 2017 international conference on computer communication and informatics (ICCCI)
3. Syedul Amin M, Jalil J, Reaz MBI (2012) Accident detection and reporting system using GPS, GPRS and GSM technology. In: 2012 international conference on informatics, electronics and vision (ICIEV)
4. Rishi R, Yede S, Kunal K, Bansode NV (2020) Automatic messaging system for vehicle tracking and accident detection. In: 2020 international conference on electronics and sustainable communication systems (ICESC). https://doi.org/10.1109/icesc48915.2020.9155836

5. Prabha C, Sunitha R, Anitha R (2014) Automatic vehicle accident detection and messaging system using GSM and GPS modem. Int J Adv Res Electr Electron Instrument Eng 3(7).https://doi.org/10.15662/ijareeie.2014.0307062
6. Dhanya S, Ameenudeen PE, Vasudev A, Benny A, Joy S (2018) Automated accident alert. In: 2018 international conference on emerging trends and innovations in engineering and technological research (ICETIETR)
7. Kalyani T, Monika S (2019) Accident detection and alert system. IJITEE 8:227–229
8. Crandall M (2019) Rapid emergency medical services response saves lives of persons injured in motor vehicle crashes. JAMA Surg 154(4):293–294. https://doi.org/10.1001/jamasurg.2018.5104
9. Fleischer PB et al (2012) Design and development of GPS/GSM based vehicle tracking and alert system for commercial intercity buses. In: 2012 IEEE 4th international conference on adaptive science and technology (ICAST). IEEE
10. EMS response time is associated with traffic accident mortality. https://www.jwatch.org/na48509/2019/02/08/ems-response-time-associated-with-traffic-accident
11. Byrne JP, Mann NC, Dai M et al (2019) Association between emergency medical service response time and motor vehicle crash mortality in the United States. JAMA Surg 154(4):286–293. https://doi.org/10.1001/jamasurg.2018.5097
12. Raj Kumar B, Jagadeesh R, Sai Kumar CH, Srinivas G, Gowri CH (2020) Accident detection and tracking system using GSM, GPS and Arduino. Int J Emerg Technol Innov Res 7(4):1773–1779. ISSN:2349-5162
13. Dalai T (2013) Emergency alert and service for automotives for India. Int J Adv Trends Comput Sci Eng (IJATCSE) 2(5):08–12
14. Agrawal A, Radhakrishnan AR, Sekhsaria A, Lakshmi P (2014) A cost effective driver assistance system with road safety features. In: 2014 international conference on embedded systems (ICES), Coimbatore, pp 211–215. https://doi.org/10.1109/EmbeddedSys.2014.6953158
15. Khalil U, Javid T, Nasir A (2017) Automatic road accident detection techniques: a brief survey. In: International symposium on wireless systems and networks (ISWSN). IEEE, pp 1–6
16. Nandaniya K et al (2014) Automatic accident alert and safety system using embedded GSM interface. Int J Comput Appl 85(6)

# Metaheuristics Algorithms for Virtual Machine Placement in Cloud Computing Environments—A Review


**Jyotsna P. Gabhane, Sunil Pathak, and Nita M. Thakare**

**Abstract** Cloud Computing provides on-demand, flexible, ubiquitous resources for clients in a virtualized environment using huge number of virtual machines (VMs). Cloud data centers don't utilize their resources fully which leads into a underutilization of resources. Virtualization offers a few exceptional highlights for cloud suppliers like saving of power consumption, load adjusting, and adaptation to internal failure, resource multiplexing. However, for improving energy proficiency and resource utilization, various strategies have been introduced such as server consolidation and different resource structuring. Among all, Virtual Machine Placement (VMP) is the most vital strides in server consolidation. Virtual Machine Placement (VMP) is an efficient mapping of VMs to Physical Machines (PMs). VMP issues go about as a non-deterministic polynomial-time hard (NP-difficult) issue and metaheuristics strategies are widely used to solve these issues with enhancing boundaries of power utilization, QoS, resource usage, etc. This paper presents an extensive review of Metaheuristics models to deal with VMP in the cloud environment.

**Keywords** Cloud computing · Data centers · Metaheuristics · Physical machines · Virtualization · Virtual machines · Virtual machine placement

## 1 Introduction

Cloud Computing is a model which shares a variety of configurable registering resourcing like systems, databases, memories, applications, software development

J. P. Gabhane (✉) · S. Pathak
Department of Computer Science and Engineering, Amity University Jaipur, Jaipur, India
e-mail: jyotspg@gmail.com

S. Pathak
e-mail: spathak@jpr.amity.edu

N. M. Thakare
Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India
e-mail: nitathakre14@gmail.com



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_28

platform, and facilities over the web [1]. The Cloud can be seen as an abstract layer on the Internet, which makes all available software and hardware resources of a data center transparent, rendering accessibility through a well-defined interface [2]. Flexibility is one of the alluring highlights of the cloud. It allows obtaining or releasing computing resources on the user's demand. It enables service providers to auto-scaled the resources to their applications which reduces the resource cost to improve the Quality of Service (QoS) [3]. Recently all storage and networking platforms moved towards the cloud in collaboration with IoT (Internet of Things), which introduces many challenges [4]. In the fast-growing cloud environment, many application providers directly host their applications on cloud resources using cloud service providers such as Amazon EC2. It allows migrating local applications to clouds, as an on-demand service [3]. Cloud computing offering infrastructure, platform, and software as pay and use services require Quality of Service (QoS) and Service-Level Agreements (SLAs) to monitor and ensure their efficient delivery. However, clients' dynamic QoS prerequisites and maintain a strategic distance from SLA infringement is a major task in cloud environments [5]. Since 2008, Cloud computing is an admired paradigm in IT Sector with three common models briefed as Software as a Service (SaaS) for applications, Infrastructure as a Service (IaaS) for hardware resources, and Platform as a Service (PaaS) for real-time application [6]. IaaS (Infrastructure as a Service) is a top-layered on-demand infrastructure resource provisioning service of cloud computing. It provides infrastructural resources like hardware requirements in virtual machines, processing powers, network bandwidth, storage, CPU, and Power [7]. IaaS users have control over data, applications, operating systems, storage, and security. Typical examples of IaaS are AT & T, GoGrid, Verizon, Amazon web services many enterprises have adopted the extremely elastic IaaS (Infrastructure as a Service) form [8]. PaaS (Platform as a Service) [9] is to shape a facilitating situation it gives stage level assets, including working framework and structures for programming advancement. The PaaS user has control over the deployment and execution of applications. It is a platform to develop, test, and deploy an application [7]. SaaS provides access to an application to the users online instead of installing or purchasing it [10]. The various deployment models of cloud computing like Private cloud, Community cloud, Public cloud, Hybrid cloud [11, 12] are used as a virtual distributed system. It has been noticed that cloud data center's energy utilization has increased due to the vast usage of cloud computing [4]. One of the ways to deal with this problem is to minimize the active machines and closing inactive servers. From the power consumption point of view, the unutilized server is highly inefficient and hence is the biggest challenge to be dealt with [13]. Virtualization technology is one of the preferable solutions to solve the power incompetence problem. Virtual machine placement (VMP) is the novel approach to increase the power competence of the cloud data center. It improves resource utilization and Return on Investment (ROI) [14]. As an emulation of a particular physical machine, a virtual machine acts as a real machine [15]. VMP improvement is a NP-hard combinatorial issue. The issue can be solved in different ways [16]. Service Level Agreements (SLA) is an agreement between the client and specialist organization, as far as service quality, responsibility, and availability. Hence, good VM allocation

is a core challenging problem nowadays to keep an optimal balance between QoS and energy consumption [3]. In the following sections, the research work proposed the idea of Cloud Computing along with Virtual Machine Placement. VM Placement comes up with the varied problem structure like VMP Optimization approaches, VMP problem formulation, and its objective functions which are discussed in detail in the following section. Existing algorithms of VMP are discussed and analyzed in the given study, finishing with the conclusion of potential answers for the virtual machine placement in the cloud environment.

## 2 Virtual Machine Placement in Cloud

To provide the services over the internet, a large number of physical devices are required to embed with various software. Every device is not fully utilized every time as the network resource usage is very unpredictable [17]. To solve such issues Virtualization is introduced for cloud environment. Virtualization is the process of partitioning the physical resources like storage, memory, network, and power of the physical machine into the multiple execution environments called Virtual Machines (VMs). Data centers are geographically distributed with VMs are located in it. The number of virtual machines available with resources such as CPU, storage space, network, and memory is effectively handled user's request in the cloud. As per the increase in the demand from the users, VM Selection plays a major role to select optimal PM to run the VM [18]. Virtual machine placement (VMP) is nothing but the allocation of proper Virtual machines at each PMs of the cloud data center. It is also called a placement plan of VM to PMs Mapping. VMP has two major objectives power-saving and delivering QoS [9]. Virtual Machines are typically portrayed by the quantity of CPU centers, CPU limit per center, RAM size, circle size, correspondence data transfer capacity and idleness between sets of VMs or a VM and the client. Requirements of VM can be stable or dynamic depending on the type of application(s) running on it. Cloud data center (DC) uses virtualization technology for serving computation, storage, and networking [19]. Virtualization has basic components like VM migration, VM fault tolerance, VM elasticity, VM consolidation, VM load balancing, and VM scheduling [20].

VMP formulation problem is classified as [21]:

- Optimization approaches
- Objective functions
- Solution techniques.

Each of them discussed details in the following sections.

## 3 VMP Optimization Approaches

Different optimization approaches discussed in the studied papers based on the various objective functions [22]. A mono-objective approach as its name indicates focused the optimization of only single objective. It is an individual streamlining of more than one target work, each in turn. As indicated by the considered articles, maximum research work follows mono target-based method for tackling the VMP issue [22]. Multi-Objective Solved As Mono-Objective Approach is a methodology where the advancement of different target capacities illuminated as a solitary target work is viewed as a multiple target way understood to be a mono-objective. To perform a right combination of the goal capacities, it requires intense information on the difficult space and as a rule; it is beyond the realm of imagination which might be expressed as a downside of this methodology [23, 24]. The approach which streamlines more than each target work, in turn, is named as Pure Multi-Objective (PMO). Even though lots of work has been done on a mono objective approach for the operational real cloud environment a PMO provides good solution [23]. Last is Many-Objective Optimization (MaOP) which focuses on more than three and up to twenty parameters. MaOPs show up generally in some certifiable applications, for example, building structure [24]. Many-objective Optimization Problems formulation is not yet projected for the VMP problem in the particular research [25].

## 4 VMP Problem Formulation/Objective Functions

As per previous research to obtain Optimal Virtual Machine Placement some important parameters like minimization of power consumption, Minimization of SLA violations, resource utilization, etc. are always need to be considered. These criteria can change from time to time depending upon the VMP Problem formulation. Various objective functions stated below are superlative for solving the VMP Problem [22, 26, 27] like High Performance, Minimizing Energy Consumption, QoS Maximization, Network Traffic Minimization, Minimizing SLA Violations, Balancing Resource Utilization, Load Balancing, VM Migration, Security, Resource Wastage Minimization, Datacenter Power Consumption Minimization, Operational Cost Minimization, Performance Maximization. Allowing for the above huge number of various objective functions, the following are Five broad classifiers [22, 27] covering almost all objectives in the following sub-sections.

- Energy Consumption Minimization
- Cost Optimization
- Network Traffic Minimization
- Resource Utilization
- Quality of Service.

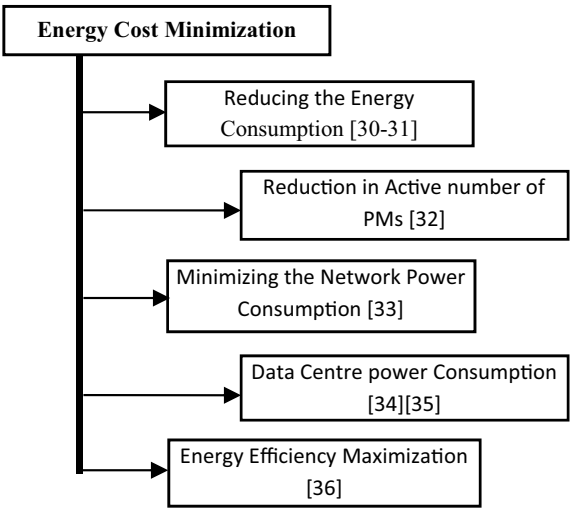



**Fig. 1** Objective functions of energy cost minimization

## *4.1 Energy Consumption Minimization*

Datacenters consume lots of power and reduction in such power consumption is a challenging task. Energy Consumption can be minimized by minimizing Active Servers [28]. This paper proposed a strategy to reduce power consumption by Maximum VM Placement to profit in revenue generation. The efficiency of energy can be increased by a reduction in VM migrations. So, the Live Migration technique is proposed in this paper for utilization of virtual machine placement by guarantee a deadlock-free resource allotment based on multi-dimensional resources. The aim is to reduce the energy consumption of datacenter [29]. Maximum studied articles focused on the main objective function of VMP which is power consumption minimization. Various target capacities sorted in this improvement group are: (1) Reducing the Energy utilization (2) Reduction in active number of PMs (3) minimizing the Network Energy utilization (4) Data Center power Consumption (5) Energy Efficiency Maximization [22] (Figs. 1 and 2).

## *4.2 Cost Optimization*

Due to the on-demand service of cloud computing, high rate is a big confront for cloud service providers. This cost includes Reducing Electricity Costs, Network Power Costs, cost of server, reduction in Connection Cost, decrease in SLA Violations as well in Operational Cost, Economical Revenue Maximization [27].

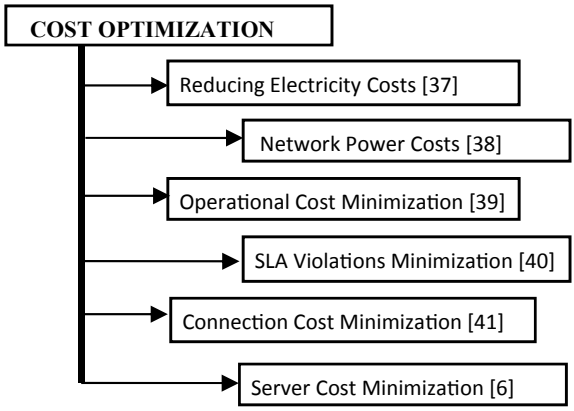



**Fig. 2** Objective functions of cost optimization

### *4.3 Network Traffic Minimization*

To enhance the performance of a data center minimizing network traffic is again a crucial parameter need to be focused on. To avoid the consumption of resources VM to VM intercommunication helps to reduce network traffic [42]. Minimization of Network traffic can be classified further by Reducing the Average Traffic Latency, reduction in the Data Transfer rate, Reduction in network traffic, Network Performance maximization [27].

### *4.4 Resource Utilization*

In data center, resources like CPU, memory, and bandwidth are important performance metrics. It aims to utilize resources by guaranteeing the QoS. In Virtualization the services provided on shared resources and utilities in Datacenters [43]. In this context, resource utilization classified further as maximizing resource usage, minimizing resource wastage, and increasing elasticity (Figs. 3 and 4).

### *4.5 Quality of Service Maximization*

Virtual Machine Placement problem can be solved using different constraints as mentioned above. One of the important constraints out of them is the Quality of Service (QoS). Reduction of the resource interference, High accessibility, Performance Improvement, Resource interference minimization, and reliability are the crucial parameters metrics to insure good QoS [27] (Fig. 5).

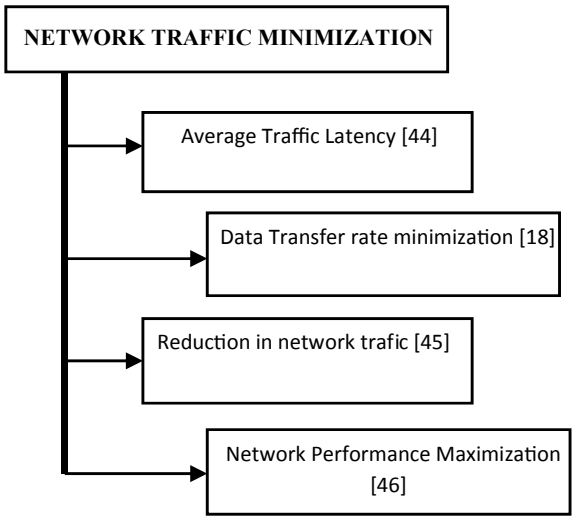



**Fig. 3** Objective functions of network traffic minimization

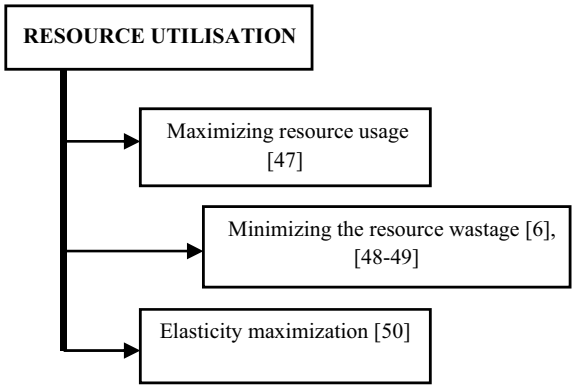



**Fig. 4** Objective functions of resource utilization

## 5 Solution Techniques

Universally, there are many approaches used to solve placement problem of Virtual machines. The major and focused techniques are deterministic algorithms, approximation algorithms, heuristics, and Metaheuristics algorithms [22]. In the following section, Metaheuristics algorithms are discussed in detail.

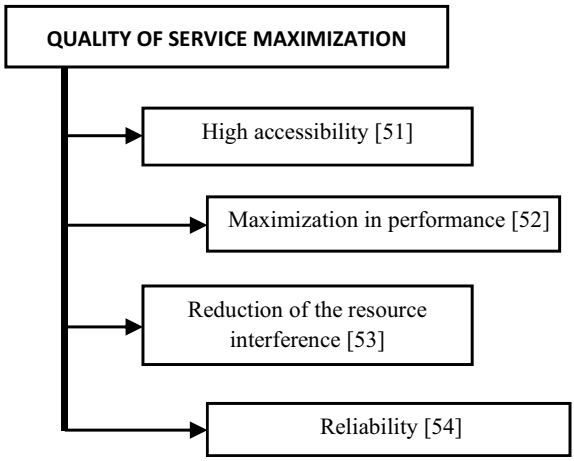



**Fig. 5** Objective functions of QoS maximization

## *5.1 Metaheuristics Algorithms*

Meta in Meta-Heuristic is a Greek word means advanced whereas heuristics defines to locate or to identify. It enhances the efficiency of the heuristic approach [55]. The VMP problem is act as or considered as a non-deterministic polynomial-time hard (NP-hard) problem. For these types of problems, it is difficult to create optimal solutions in a short period. Metaheuristics techniques are one of the promising solutions which can provide nearby perfect and feasible solutions within a preferable instance. Because of efficiency of solving large and complex problems, Metaheuristics admired from last few years. Many Metaheuristics algorithms have been used to solve the VMP difficulties in terms of power, resources, network traffic, and cost optimization [56]. The following section discusses more in detail about Metaheuristics approaches.

**Ant Colony Optimization (ACO)**

A VMP approach used the concept of the Ant Colony algorithm was projected in [57] to enhance power utilization and load balancing of the system. Idle resources and power consumption model was proposed as multi-objective technique and the utilized parameters of virtual machines. For the performance assessment of the method the comparison is done with the traditional genetic algorithm-based MGGA model. Their experimental results indicate that the rate of convergence is increased by 16%, and the average energy consumption is reduced by 18%. This paper [58] proposed an ant colony optimization method for the consolidation of virtual machine (VM) in cloud environment. Some VM migrations plans were proposed for underutilized PMs and are used to reduce the number of PMs. They have considered multi objectives for minimizing VM migration. The researcher claimed that when their algorithm gets compared with existing algorithms, it outperformed in terms of VM Migration, the number of unrestricted PMs as well packing effectiveness. A power proficient

OEMACS for VMP in cloud environment was proposed by [59]. To achieve optimal VM deployment number of dynamic servers needs to be reduced. It avoids possible wastage of computation. An ACS-based approach is proposed coupled with order exchange and migration (OEM) so named as an OEMACS. This scheme focused on reducing working servers with good utilization of energy, consumption of resource, and proper balancing. Finally, they compared the performance of the OEMACS with heuristic and other evolutionary methods, which generally outperforms in resource characteristics, and power optimization than other algorithms. The author [36] has formulated VMs to PMs in a data center as an inhibited problem which uses data from PM to VM to reduce down the full energy utilization of working PMs. The focus is on reduction of power utilization of cloud centers. They proposed an Ant colony scheme implanted through a heuristic approach. The experimental results compared with the two existing ACO methods, ACOVMP and PAVM as well as First-Fit-Decreasing (FFD) algorithm. The authors claimed that their proposed algorithm has an improvement in energy efficiency by 5% (Table 1).

**Table 1** The summary of ant colony optimization

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [57] | To Optimized Energy consumption and load Balancing | Energy aware | ACO | The proposed algorithm of this paper acts as an energy-saving model |
| [58] | To reduce overprovisioning of physical machines (PMs) | Energy-aware | ACO | The proposed algorithm outperformed in terms of VM Migration, the number of released PMs, and packing efficiency |
| [59] | To minimize the number of active servers, and to improve the resource utilization, load balancing, and reducing power consumption | Resource aware Energy aware | OEMACS | The proposed algorithm outperformed for resource characteristics, and power optimization |
| [36] | To minimize the power consumption of a data center | Energy aware | ACO | The proposed algorithm gave an improvement in energy efficiency by 5%. It also showed a 10% improvement in the new ACS |

**Particle Swarm Optimization (PSO)**

In 2015 [35] authors have proposed the power-aware Particle Swarm Optimization method for the virtual placement of virtual machine to decrease utilization of power in the cloud data centers. The projected approach compared with the non-power-aware algorithm and proved to be superlative. This paper [60] presented a multi-objective optimization model to address the VM Migration problem. VM Migration can generally affect the performance in terms of load balancing, fault tolerance, utilization of energy, and utilization of resource. The proposed algorithm Fuzzy Particle Swarm Optimization (PSO) used fuzzy logic for regulation inertia mass along with conventional particle swarm optimization to resolve the mapping issues. This paper proposed [48] an improved PSO solution for the heterogeneous data center. The aim of the given approach is to fit with proper placement of virtual machine in data hubs with optimal energy consumption. This paper mainly focused on Energy-efficiency of green computing. The proposed algorithm compared the performance with First fit, Best fit, and worst fit strategies and claimed to do a better use of the resource. A resource aware VMP approach stand on the concept of PSO meta-heuristics was proposed in [30] to increase packing effectiveness while reducing power utilization. It was a multi-objective version that considers energy and resource utilization. The VMP Formulation was presented to improve the packing effectiveness and so reduction in power utilization. The simulation outcome showed that the planned technique exhibits excellent performance. As the drawback, the proposed system works only for the static workload (Table 2).

**Biogeography-Based Optimization (BBO)**

In this literature [6] a novel approach named VMPMBBO projected to discover the best possible placement of virtual machines which reduces wastage of resources and utilizes proper consumption of energy. The author claimed that the proposed technique gives the primary approach that uses Biogeography-Based Optimization (BBO) and composite systems for solving the VMP problem. Finally, it proved that VMPMBBO has recovered authentic properties with more efficient and robust environment. Experiments were conducted on both real-world data and synthetic data from research. Flexibility and scalability of VMPMBBO also proved via various practical outcomes. In 2016, a multi-objective VMP optimization problem [61] with an improved BBO algorithm is proposed named IMBBO (Improved BBO). In the major contributions of the given article, the focused is on the VMC problem a multi-objective task. It includes parameters like minimizing server power consumption, proper loading balancing, and reducing migration overhead. Experimental results based on synthetic datasets and real VMs running data, the proposed IMBBO compared with Gravitational Search Algorithm (GSA) which showed improvement in performance (Table 3).

**Genetic Algorithm (GA)**

A procedure proposed [31] for controlling power consumption with the active Voltage rate range system is included in the optimization system and Non-dominated Sorting

**Table 2** The summary of particle swarm optimization

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [35] | To reduce the power consumption of the data centers | Energy-aware | PSO | A model is proposed to minimize in power consumption of a data centers |
| [60] | To reduce energy consumption and to maximize the resource utilization | Energy-aware Resource aware Cost aware | PSO | The proposed algorithm improves the efficiency of conventional PSO by using the fuzzy logic system and used to solve the optimization problem. The future research direction is to consider the selection of proper PMs to hosts the VMs |
| [48] | To fit with virtual machine placement in data centers with minimum power consumption | Energy aware Resource aware | Improved PSO | The proposed algorithm uses resource wastage and energy consumption efficiently. The extension of this work can make use of better energy-efficiency models |
| [30] | To improve packing efficiency while reduction in energy consumption | Energy aware Resource aware | PSO | The VMP Formulation model was proposed to maximize the packing efficiency and reducing the energy consumption. The drawback of the proposed system is that it works for static workload only |

Genetic Algorithm (NSGA-II) used for a set of non-domination solutions. The primary goal of this method is to minimizing the energy consumption as well as execution time. To predict Virtual Machine based on so predicting Artificial Neural Network (ANN) is applicable for envisage the virtual machines performance depends on the tasks given as well resource allocations. The results showed that effective minimization of energy consumption by including DVFS in the optimization approach. The VMP bi-objective combinatorial optimization approach was proposed [49]. An

**Table 3** The summary of biogeography-based optimization

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [6] | To find the optimal VM placements for minimizing power consumption and resource wastage | Energy and resource-aware | VMPMBBO | The proposed algorithm gives optimal VM placements which simultaneously minimizes both the resource wastage and energy consumption. In future author will more focus on the Parallelization of the algorithms |
| [61] | To improve the classical BBO algorithm | Energy and resource aware | Improved BBO | The proposed strategy minimized server power consumption with proper loading balancing and reduced migration overhead. The future will more focus on the parallelization of IMBBO concerning the re-configuration migration and mutation model in the real |

efficient decision-making structure of genetic algorithm and Bernoulli simulation (GAVMP-BO) is proposed for multi objectives which are used to reduce dynamic PMs. The main objective of the research is reduction of unbalanced resources in dynamic PMs. The proposed algorithm was compared to six different competing algorithms. Results show that the proposed work is superlative and outperformance than all frameworks (Table 4).

**Simulated Annealing (SA)**

In 2015, a Simulated Annealing based algorithm was proposed [62] for improving the performance in terms of cost. A novel approach was introduced to solve the VM consolidation problems by estimating the cost of possible VM migrations. Network traffic and different network topology parameters for consolidation decision left as future work part in this literature. Getting optimal VMP environment is little bit cumbersome method in terms of time. To resolve the issue, this research [63] proposed an improved simulated annealing algorithm (ISA) for VMP. For servers, resource utilization and for VM dynamic models were introduced. The threshold value defined for the annealing process in ISA and the allocation weight vector

**Table 4** The summary of genetic algorithms

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [31] | To reduce the makespan and energy power consumption of the cloud services | Energy aware Resource aware Cost aware | NSGA II with ANN | The results of proposed approach showed the effective minimization of energy consumption |
| [49] | To reduce power consumption with the operational costs and resource wastage | Energy aware Resource aware Cost aware | GAVMP-BO | The proposed work is superlative and outperformance than all frameworks |

was computed. Comparative analysis showed that the approach got a desired output within a smaller amount time. The mentioned approach improved resource utilization with proper load balancing. In this paper [28] a method called maximum VM placement with least utilization of energy (MVMP) was proposed to increase the revenue of a Cloud Service Provider (CSP). The given method aimed reduction in power cost as well increase the revenue. The proposed algorithm compared with different techniques. MVMP has improved performance than all approaches in power consumption, execution time, and the number of servers used and it is scalable too (Table 5).

**Artificial Bee Colony (ABC)**

In a Heterogeneous Cloud environment for the reduction in the make-span this paper [64] proposed Heuristic task scheduling combined with the Artificial Bee Colony algorithm (HABC) for the virtual machine. It is a task scheduling approach with proper load balancing. The two major goals of the given method were to maximize productivity by balancing workload and to reduce the total amount of time. The experimental outcomes proved that the given approach improved the effectiveness in scheduling the task as well as load balancing of virtual machines in given environment. Energy-related dynamic VM Consolidation approach (EC-VMC) was proposed [65]. The experimental outcome indicated that the proposed algorithm exceptional in provisions of reducing utilization of power, VM immigration with effective QoS (Table 6).

**Firefly Algorithm (FA)**

Many research focused on Virtual machine placement parameters like energy consumption, resource utilization, and network traffic. Very few works considered security and privacy issues. This [66] paper proposed a secure and efficient VMP framework by incorporating a proper constraints-based multi-objective model using the Discrete Firefly algorithm (DFA-VMP).To calculate the security and efficiency of the system authors have specially defined the security manifestation, with calculated power consumption as well focus on reduction in loss of resources. The outcomes express that this methodology adequately decreased the likelihood of co-occupant

**Table 5** The summary of simulated annealing

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [62] | To improve the performance of the system in terms of cost | Energy aware<br>Resource aware<br>Cost aware | SA | A novel method was introduced to solve the VM consolidation problems but the parameters like Network traffic and different network topology left as future work part in this literature |
| [63] | To improve the resource utilization rate, allocation of resources with steady and capable working of data center | Resource aware | ISA | The proposed method converges faster and improved utilization rate of resources compared with traditional ones. However, some factors such as response time are not considered in this research |
| [28] | To maximize the revenue and minimize the power cost | Energy aware<br>Cost aware | MVMP | MVMP performs better than all approach in terms of power consumption, execution time and number of servers used and it is scalable too. Dynamic VM Placement is left for future work |

**Table 6** The summary of artificial bee colony

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [64] | In order to stabilize the workload the processing time as well total execution time gets reduced | Resource Aware | HABC | This algorithm improved the efficiency in task scheduling as well in load balancing |
| [65] | To proposed energy-aware VM consolidation model | Energy aware | EC-VMC | The proposed algorithm exceptional for reduction in power consumption, VM migrations, and improving QoS |

**Table 7** The summary of firefly algorithm

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [66] | To deal with the issues of IaaS platform like side channel attack | Security Aware | DFA-VMP | The related security indices, measured power consumption, and loss of resource at the data center are considered to measure the system security and efficiency |

in a similar hub of VMs of pernicious inhabitants with the objective virtual machine and furthermore diminished power utilization and loss of resources at the server.

**Tabu Search (TS)**

To reduce power consumption and reduction in CO2 emission a very efficient Tabu search heuristic technique [33] was proposed. It solved the Mixed Integer Linear Programming (MILP) model. To handle online and dynamic demands Tabu search a Metaheuristics approach was implemented. The result stated that reduction in Communication delays up to 6 times. CO2 emissions reduced up to 60 times and 30% of power savings [67]. The key idea of the proposed work was to achieve energy-efficient VM Allocation plans utilizing the hypothesis of heartiness and determine a strong Mixed Integer Linear programming (MILP) detailing. The goal of the calculation is to build up a VM-to-PM designation which lessens the power utilization, the quantity of relocations and guarantees strength (Table 7).

Tabu Search is used as base which incorporates a greedy approach. It showed an effective and stable performance than various Metaheuristics algorithms. Various local search algorithms can utilize this algorithm. The exploration guaranteed that the developments indicated the vigorous heuristic can figure most ideal arrangements and beats the notable VM Workload Consolidation Strategy (WCD) heuristic (Table 8).

**Cuckoo Search (CS)**

In 2018, [68] an Enhanced Cuckoo Search (ECS) was proposed to optimize utilization of power in VMP. The focus of work is to how to narrow down utilization of power along with resources in the cloud data center. The result of the proposed algorithm compared with the existing Genetic Algorithm (GA,) Optimized Firefly Search Algorithm (OFS), and Ant Colony (AC) algorithm and proved that it consumes less energy without violating SLA. For evaluating the performance of the proposed ECS work showed better energy metric with VM selection policies like Minimum Migration Time (MMT) and Minimum Utilization (MU) (Table 9).

**Table 8** The summary of tabu search

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [33] | To reduce energy consumption and reduction in $CO_2$ emission | Energy aware Cost aware | Tabu Search | The result stated that reduction in communication delay up to 6 times. $CO_2$ emissions reduced up to 60 times and 30% of power savings |
| [67] | The objective of the algorithm is to establish a VM-to-PM allocation which reduces the energy consumption, the number of migrations and ensured robustness | Energy aware | Γ-robust heuristic (Tabu search) | The evolutions showed the robust heuristic can calculate optimal solutions and outperforms the well known VM workload consolidation strategy (WCD) heuristic |

**Table 9** The summary of cuckoo search

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [68] | Reduction in energy consumption and resource utilization | Energy aware Resource aware | Enhanced cuckoo search (ECS) | A real heterogeneous environment is provided with some performance evaluation metrics left as a future work |

### Memetic Algorithm (MA)

This work [25] presented a Many-Objective Optimization Structure (MaVMP) for solving static means offline problems of virtual machine placement. An interactive memetic algorithm is proposed to solve the formulated MaVMP problem, considering context of many objectives. For the formulation of a many-objective VMP problem objectives like power, cost, security, quality of service, network load balancing are considered. Experimental results proved that it is an effective and capable solution to solve the problem. In [34] an improved genetic algorithm as memetic grouping used for cost-efficient VM placement, namely MGGAVP is proposed in the multi-cloud environment. The research presented an optimization model in the heterogeneous environment with efficient virtual machine placement. The parameters on-demand real-time and topographical situation of apportioned

**Table 10** The summary of memetic algorithm

| References | Objectives | Problem addressed | Algorithms | Comments |
|---|---|---|---|---|
| [25] | To solve offline VMP problem | Energy aware<br>Cost aware<br>Resource aware | MaVMP | Experimental results proved that it is effective and capable solution to solve the offline problems of VMP |
| [34] | To provide energy and cost-efficient VM placement | Energy aware<br>Cost aware | MGGAVP | The proposed algorithm compared and proved better performance in energy saving |

resources were considered as the objective of diminishing power cost. The proposed model reduced the number of working PMs as well power consumption of network in distributed data centers. The proposed algorithm compared with CRA-DP, E-aware, and CTPS and proved significantly better performance (Table 10).

## 6 Discussion

A summary of the optimization methods from the various research articles is demonstrated in different Tables. The fundamental goal of this literature is to examine and checking on a preferred study concerning Virtual Machine Placement in distributed environment utilizing Metaheuristics Algorithms. Energy utilization and resource usage are priority goals to be achieved in the VMP as multiobjective issues. VMP is characterized into certain unique classes including power, resources, security, and cost. According to Fig. 6, all mentioned parameters are regularly studied in research.

Above figure shows that researchers regularly tended to and assessed the VMP Environment using energy consumption as primary factor. In any case, other significant destinations should be viewed as important parameters of VMP like security, dynamic load balancing, network traffic management, resource utilization. The hybrid Metaheuristics approaches may get benefits from different algorithms by considering various parameters in multi-objective way. Hybrid Metaheuristics algorithms can be the possible solution that will help to enhance the performance of VMP. From the reviews, it is observed that Hybrid Metaheuristics algorithms have once in a while been utilized in the VMP. Therefore, there is a future scope to consider multiobjective hybrid metaheuristics algorithms for virtual machine placement.

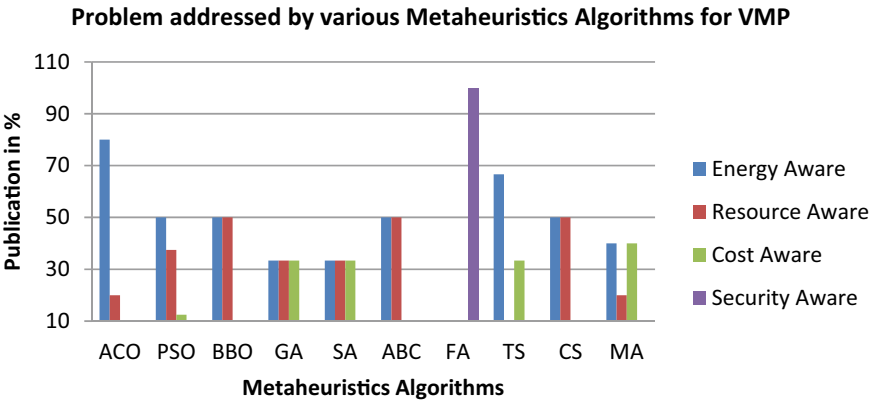



**Fig. 6** Problem addressed by various metaheuristics algorithms for VMP

## 7 Conclusion

Evaluation of Metaheuristics methods for virtual machine placement in the cloud Environment is analyzed in this research work. The investigation of VMP calculations looked at the different advancement strategies and various target capacities. From the analysis, it is observed that most of the research works are introduced to limit the power utilization of data centers. A few researchers have additionally tended to issues identified with the execution and utilization of resources. The primary aim is to reduce the power utilization of DCs without degrading execution or violating SLA limitations. Virtualization has empowered cloud data centers to get cost-effective, flexible, and adjustable to have various undertaking administrations and applications at various scales. With the advancement of virtualization innovation, methods to deploy VMs in the cloud are a hot research subject. Due to the extending size of cloud data centers, the amount of customers and virtual machines (VMs) increases rapidly. The solicitations of clients are engaged by VMs living on physical servers. The dynamic development of internet providers brings about asymmetrical system resources. Resource management is a significant factor for the exhibition of a cloud. Different methods are utilized to deal with the resources of a cloud efficiently. It is a great initiative to think about Multi-objective Virtual machine placement for new structured nature propelled Metaheuristics algorithms as future work objectives considering the parameters like network traffic, load balancing, power utilization, and resource usage. Virtual machine placement can be unraveled for various destinations. So hybridizing advantages of such algorithms can likewise give another vision towards the Virtual machine placement issue.

## References


1. Mell P, Grance T (2011) The NIST definition of cloud computing
2. Endo PT et al (2016) High availability in clouds: systematic review and research challenges. J Cloud Comput 5(1):16
3. Qu C, Calheiros RN, Buyya R (2018) Auto-scaling web applications in clouds: a taxonomy and survey. ACM Comput Surv (CSUR) 51(4):1–33
4. Abreu DP et al (2019) A comparative analysis of simulators for the cloud to fog continuum. Simul Modell Pract Theor:102029
5. Singh S, Chana I (2015) QoS-aware autonomic resource management in cloud computing: a systematic review. ACM Computing Surveys (CSUR) 48(3):1–46
6. Zheng Q et al (2016) Virtual machine consolidated placement based on multi-objective biogeography-based optimization. Future Gener Computer Syst 54:95–122
7. Da Cunha Rodrigues G et al (2016) Monitoring of cloud computing environments: concepts, solutions, trends, and future directions. In: Proceedings of the 31st annual ACM symposium on applied computing
8. Zhang Qi, Cheng Lu, Boutaba R (2010) Cloud computing: state-of-the-art and research challenges. J Internet Serv Appl 1(1):7–18
9. Usmani Z, Singh S (2016) A survey of virtual machine placement techniques in a cloud data center. Proc Comput Sci 78:491–498
10. Fatima A et al (389) Virtual machine placement via bin packing in cloud data centers. Electronics 7(12):389
11. Saber T et al (2018) VM reassignment in hybrid clouds for large decentralised companies: a multi-objective challenge. Future Gener Comput Syst 79:751–764
12. Fatima A et al (2019) An enhanced multi-objective gray wolf optimization for virtual machine placement in cloud data centers. Electronics 8(2):218
13. Chowdhury MR, Mahmud MR, Rahman RM (2015) Implementation and performance analysis of various VM placement strategies in CloudSim. J Cloud Comput 4(1):20
14. Beloglazov A, Buyya R (2012) Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurr Comput Practice Experience 24(13):1397–1420
15. Arab HB (2017) Virtual machines live migration. PDF document. Available at 2017
16. Shirvani MH, Rahmani AM, Sahafi A (2020) A survey study on virtual machine migration and server consolidation techniques in DVFS-enabled cloud datacenter: taxonomy and challenges. J King Saud Univ Comput Inf Sci 32(3):267–286
17. Dwiardhika D, Tachibana T (2019) Virtual network embedding based on security level with VNF placement. Secur Commun Netw
18. Shahapure NH, Jayarekha P (2018) Distance and traffic based virtual machine migration for scalability in cloud computing. Proc Comput Sci 132:728–737
19. Kamaludin H et al (2020) Implementing virtual machine: a performance evaluation. In: International conference on soft computing and data mining. Springer, Cham
20. Gill SS, Buyya R (2018) A taxonomy and future directions for sustainable cloud computing: 360 degree view. ACM Comput Surv (CSUR) 51(5):1–33
21. López-Pires F, Barán B (2017) Cloud computing resource allocation taxonomies. IJCC 6(3):238–264
22. Lopez-Pires F, Baran B (2015) Virtual machine placement literature review. arXiv:1506.01509
23. Vahed D, Nasim MG-A, Souri A (2019) Multiobjective virtual machine placement mechanisms using nature-inspired metaheuristic algorithms in cloud environments: a comprehensive review. Int J Commun Syst 32(14):e4068
24. Li B et al (2015) Many-objective evolutionary algorithms: a survey. ACM Comput Surv (CSUR) 48(1):1–35
25. López-Pires F, Barán B (2017) Many-objective virtual machine placement. J Grid Comput 15(2):161–176

26. Divya BP, Prakash P, Vamsee Krishna Kiran M (2017) Virtual machine placement strategies in cloud computing. 2017 Innov Power Adv Comput Technol (i-PACT)
27. Attaoui W, Sabir E (2018) Multi-criteria virtual machine placement in cloud computing environments: a literature review. arXiv:1802.05113
28. Addya SK et al (2017) Simulated annealing based VM placement strategy to maximize the profit for cloud service providers. Eng Sci Technol Int J 20(4):1249–1259
29. Choudhary A, Rana S, Matahai KJ (2016) A critical analysis of energy efficient virtual machine placement techniques and its optimization in a cloud computing environment. Proc Comput Sci 78(C):132–138
30. Braiki K, Youssef H (2018) Multi-objective virtual machine placement algorithm based on particle swarm optimization. In: 2018 14th international wireless communications and mobile computing conference (IWCMC). IEEE
31. Sathya SA, GaneshKumar P (2018) Multi-objective task scheduling to minimize energy consumption and makespan of cloud computing using NSGA-II. J Netw Syst Manage 26(2):463–485
32. Al-Moalmi A et al (2019) Optimal virtual machine placement based on grey wolf optimization. Electron 8(3):283
33. Larumbe F, Sanso B (2016) Green cloud broker: On-line dynamic virtual machine placement across multiple cloud providers. In: 2016 5th IEEE international conference on cloud networking (Cloudnet). IEEE
34. Rashida SY et al (2019) A memetic grouping genetic algorithm for cost efficient VM placement in multi-cloud environment. Cluster Comput:1–40
35. Aruna P, Vasantha S (2015) A particle swarm optimization algorithm for power-aware virtual machine allocation. In: 2015 6th international conference on computing, communication and networking technologies (ICCCNT). IEEE
36. Alharbi F et al (2019) An ant colony system for energy-efficient dynamic virtual machine placement in data centers. Expert Syst Appl 120:228–238
37. Le K et al (2011) Reducing electricity cost through virtual machine placement in high performance computing clouds. Proceedings of 2011 international conference for high performance computing, networking, storage and analysis
38. Fang W et al (2013) VMPlanner: optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers. Comput Netw 57(1):179–196
39. Lakkakorpi J et al (2016) Minimizing delays in mobile networks: With dynamic gateway placement and active queue management. In: 2016 wireless days (WD). IEEE
40. Zhang X, Yue Q, He Z (2014) Dynamic energy-efficient virtual machine placement optimization for virtualized clouds, vol II. In: Proceedings of the 2013 international conference on electrical and information technologies for rail transportation (EITRT2013). Springer, Berlin, Heidelberg
41. Fukunaga T, Hirahara S, Yoshikawa H (2015) Virtual machine placement for minimizing connection cost in data center networks. In: 2015 IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE
42. Ihara D, López-Pires F, Barán B (2015) Many-objective virtual machine placement for dynamic environments. In: 2015 IEEE/ACM 8th international conference on utility and cloud computing (UCC). IEEE
43. Chen T et al (2018) Improving resource utilization via virtual machine placement in data center networks. Mobile Netw Appl 23(2):227–238
44. Kuo J-J, Yang H-H, Tsai M-J (2014) Optimal approximation algorithm of virtual machine placement for data latency minimization in cloud systems. In: IEEE INFOCOM 2014-IEEE conference on computer communications. IEEE
45. Ilkhechi AR, Korpeoglu I, Ulusoy Ö (2015) Network-aware virtual machine placement in cloud data centers with multiple traffic-intensive components. Comput Netw 91:508–527
46. Taleb T, Bagaa M, Ksentini A (2015) User mobility-aware virtual network function placement for virtual 5G network infrastructure. In: 2015 IEEE international conference on communications (ICC). IEEE

47. Hieu NT, Di Francesco M, Jääski AY (2014) A virtual machine placement algorithm for balanced resource utilization in cloud data centers. In: 2014 IEEE 7th international conference on cloud computing. IEEE
48. Abdessamia F et al (2017) An improved particle swarm optimization for energy-efficiency virtual machine placement. In: 2017 international conference on cloud computing research and innovation (ICCCRI). IEEE
49. Riahi M, Krichen S (2018) A multi-objective decision support framework for virtual machine placement in cloud data centers: a real case study. J Supercomput 74(7):2984–3015
50. Ortigoza J, López-Pires F, Barán B (2016) Dynamic environments for virtual machine placement considering elasticity and overbooking. arXiv:1601.01881
51. Herker S et al (2015) Data-center architecture impacts on virtualized network functions service chain embedding with high availability requirements. 2015 IEEE Globecom workshops (GC Wkshps). IEEE
52. Tordsson J et al (2012) Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. Future Gener Comput Syst 28(2):358–367
53. Kim G et al (2012) Virtual machines placement for network isolation in clouds. In: Proceedings of the 2012 ACM research in applied computation symposium
54. Zhou A et al (2016) Cloud service reliability enhancement via virtual machine placement optimization. IEEE Trans Serv Comput 10(6):902–913
55. Ullah A et al (2019) Artificial bee colony algorithm used for load balancing in cloud computing
56. Alboaneen DA, Tianfield H, Zhang Y (2016) Metaheuristic approaches to virtual machine placement in cloud computing: a review. 2016 15th international symposium on parallel and distributed computing (ISPDC). IEEE
57. Zhang L et al (2016) Towards energy efficient cloud: an optimized ant colony model for virtual machine placement. J Commun Inf Netw 1(4):116–132
58. Ashraf A, Porres I (2018) Multi-objective dynamic virtual machine consolidation in the cloud using ant colony system. Int J Parallel Emergent Distrib Syst 33(1):103–120
59. Liu X-F et al (2016) An energy efficient ant colony system for virtual machine placement in cloud computing. IEEE Trans Evol Comput 22(1):113–128
60. Ramezani F, Naderpour M, Lu J (2016) A multi-objective optimization model for virtual machine mapping in cloud data centres. In: 2016 IEEE international conference on fuzzy systems (FUZZ-IEEE). IEEE
61. Shi K et al (2016) Multi-objective biogeography-based method to optimize virtual machine consolidation. SEKE
62. Marotta A, Avallone S (2015) A simulated annealing based approach for power efficient virtual machines consolidation. In: 2015 IEEE 8th international conference on cloud computing. IEEE
63. Su N et al (2016) Research on virtual machine placement in the cloud based on improved simulated annealing algorithm. In: 2016 world automation congress (WAC). IEEE
64. Kimpan W, Kruekaew B (2016) Heuristic task scheduling with artificial bee colony algorithm for virtual machines. In: 2016 joint 8th international conference on soft computing and intelligent systems (SCIS) and 17th international symposium on advanced intelligent systems (ISIS). IEEE
65. Li Z et al (2018) Energy-aware and multi-resource overload probability constraint-based virtual machine dynamic consolidation method. Future Gener Comput Syst 80:139–156
66. Ding W et al (2018) DFA-VMP: An efficient and secure virtual machine placement strategy under cloud environment. Peer-to-Peer Netw Appl 11(2):318–333
67. Nasim R, Kassler AJ (2017) A robust Tabu Search heuristic for VM consolidation under demand uncertainty in virtualized datacenters. In: 2017 17th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGRID). IEEE
68. Barlaskar E, Singh YJ, Issac B (2018) Enhanced cuckoo search algorithm for virtual machine placement in cloud data centres. Int J Grid Utility Comput 9(1):1–17

# Prostate Image Segmentation Using Ant Colony Optimization-Boundary Complete Recurrent Neural Network (ACO-BCRNN)


**J. Ramesh and R. Manavalan**

**Abstract** Image Segmentation plays an indispensable role in Computer Aided Diagnosis system to extract the region of interest for diagnosis. Transrectal Ultrasonography (TRUS) is the best imaging technique used mostly by physicians to separate the prostate region from tissues around it and the same used for abnormality detection. The extraction of a region of interest from TRUS images is complex as it contains speckle noise, tracker, low dissimilarity, the fuzzy region between object and background, and also irregular form and size. To resolve these problems, a novel approach Ant Colony Optimization with Boundary Complete Recurrent Neural Network (BCRNN) is aimed to take out an accurate prostate region from the TRUS images. This method comprises two stages such as pre-processing and segmentation. In the beginning stage, Ant Colony Optimization (ACO) method is adopted to eradicate speckle noise from the TRUS prostate image. And, in the second stage Boundary Complete Recurrent Neural Network (BCRNN) method along with shape prioritize and multi-view fusion is employed to draw out the rigorous shape without boundary loss of the prostate. BCRNN comprises of three modules. Firstly, images are serialized into dynamic sequences and Recurrent Neural Network (RNNs) is applied to obtain shape prior to an image. Secondly, multi-view fusion approach is embedded to merge the shape predictions attained from various perspectives. Finally, to refine the details of the shape prediction map, the RNN core method is inserted into a multiscale auto-context. The proposed method is assessed using statistical performance metrics such as Dice Similarity Coefficient (DICE), Jaccard index (JI), Precision, Recall, and Accuracy. From the result analysis, it is found that ACO with BCRNN successfully extracts Region of Interest (ROI) from the image with an accurate boundary that is used for diagnosis. The experiment results clearly revealed that the performance of ACO with BCRNN is superior to BCRNN.

**Keywords** Ultrasound prostate image · Speckle noise · Image segmentation · ACO · BCRNN

J. Ramesh (✉) · R. Manavalan
K.S. Rangasamy College of Arts and Science (Autonomous), Tiruchengode, India

Department of Computer Science, Arignar Anna College Arts College, Villupuram, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_29

## 1 Introduction

Prostate Cancer (PC) is a serious disease found exclusively in the reproductive system of men. Prostate cancer is formed by the continuous deposition of protein in the prostate gland. It is the second leading cause of cancer death in the United States. The American cancer society predicted that there will be 191,930 new cases of prostate cancer with a fatality rate of 19.1% in the year 2019 [1]. Ultrasound imaging is the most widely used technology for prostate biopsy. The segmentation technique is used to segregate an image into meaningful parts with the same properties and features. Relies on the application and image attributes, different segmentation techniques are used for the prostate image. In the medical image segmentation method, the image is categorized into non-covering pixels areas that identify with eminent anatomical structures, such as, malignant growth or cyst. Segmentation techniques are broadly split into three types such as supervised, unsupervised, and interactive. In the supervised method, the manually labeled training data are utilized to extract the ROI [2–5]. In an unsupervised method, separations of an object are done without manual intervention [6]. In the iterative method, images are segmented by a mixture of human experts and machine intelligence. The interactive segmentation plays an important role in various applications like managing tumors identification, assessing tissue volumes, the diagnosis system helped medical procedure, and diagnosing sicknesses. Transrectal Ultrasound (TRUS) is used to scan the prostate and its near tissues. It is also known as Prostate Sonogram or Transrectal ultrasound.

Transrectal Ultrasound (TRUS) is a cost-effective, portable, and quick medical image method offering an instinctive portrayal of the crucial anatomic structure of the body. The reasons for the difficulties in finding the accurate edges in the TRUS image are destitute contrast, speckle noise, and loss of boundary. Hence, the automatic segmentation method is necessary to extract the prostate from the image for diagnosis. Many automated methods have been implemented for the segmentation of prostate from the TRUS image. However, the methods have remained a challenging task. In this research paper, the novel method called automatic ultrasound prostate image segmentation technique using Ant Colony Optimization with BCRNN is proposed. An overview of the proposed framework for TRUS image segmentation is shown in Fig. 1.

The rest of the paper is organized as follows: Sect. 2 briefly describes the materials and methods for prostate segmentation. The detailed description of the proposed methodology is discussed in Sect. 3. In Sect. 4, the experimental analysis using evaluation metrics is presented and the research work is concluded in Sect. 5 with future scope.

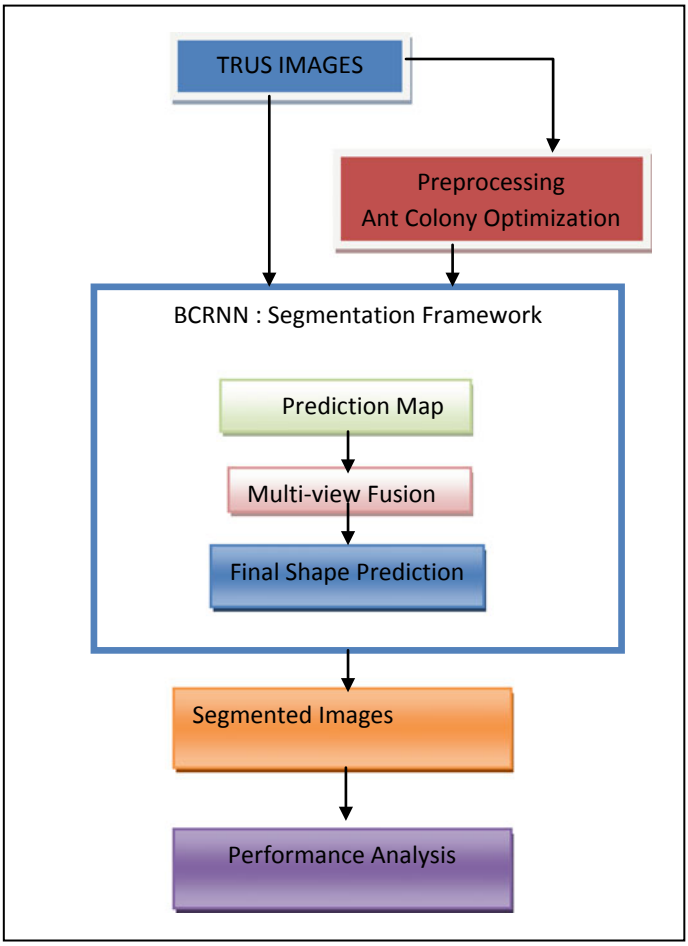



**Fig. 1** An overview of proposed framework

## 2 Materials and Methods

Generally, boundary detection is a very difficult task in TRUS prostate images due to a very low signal-to-noise ratio. The traditional methods of edge detection struggle to determine the exact boundary region. Recently, various image segmentation approaches have been applied to extract the prostate region, and the same are explained hereunder. Shoo et al. [7] furnished an overview of boundary detection methods used for segmentation. Wu et al. [8] applied Genetic Algorithms for acquiring region of interest from prostate image. Bernard Chiu et al. [9] introduced Dyadic Wavelet Transform and Discrete Dynamic Contour for prostate image segmentation. Cruz-Aceves et al. proposed Multiple Active Contours driven by

Particle Swarm Optimization for Cardiac Medical Image Segmentation [10]. Non-parametric multi-level image segmentation technique was developed by Otsu et al. [11] and it was modified by Kapur et al. [12]. Many of the image segmentation methods are not acquiring the boundaries accurately, since, it is based on histogram techniques. Luca and Termini [13] introduced a fuzzy partition technique for image segmentation and Bloch et al. [14] implemented the applications of fuzzy spatial relationship in image process. Zhao et al. [15] applied a multi-level approach by defining three membership functions for three levels thresholding such as dark, medium, and bright. Tao et al. [16] suggested a three-level fuzzy entropy-based method for image segmentation. Cao et al. [17] applied global optimization method Genetic Algorithm (GA) to attain the maximum entropy value. Cootes et al. [18] implemented the Active Shape Model (ASM) to achieve both the shape and information volume of the prostate. Shen et al. [19] proposed ASM with Gabor descriptor for prostate ultrasound image segmentation. Van Ginneken et al. [20] introduced ASM along with Gaussian to extract the image boundary. Rogers et al. [21] is proposed Robust Active Shape Model to reject displacement outliers in the image and the same were also introduced for prostate ultrasound image by Santiago et al. [22]. Ascertaining the low-rank property of similar shapes, an additional reliability constraint is expanded to acquire a shape model without boundary deficiency in ultrasound images proposed by Zhou et al. [23]. Yan et al. [24] introduced the Partial Active Shape Model to find the missing boundaries of the TRUS prostate image. Yang et al. [25] proposed Boundary Complete Recurrent Neural Network (BCRNN) for obtaining the accurate boundary and shape of the prostate ultrasound image (raw image) without any loss. In this research, Ant Colony Optimization (ACO) is hybridized with BCRNN to acquire region and boundary accurately from prostate image. The prostate region and segmentation techniques are explained hereunder.

## 3 Proposed Methodology

In the proposed novel approach is introduced to segment the prostate region from ultrasound image. The segmentation framework is illustrated in Fig. 1. Each step is explained in the following subsections. The method consists of two stages. (i) ACO method is used to remove the speckle noise from prostate image and (ii) BCRNN method is implemented to gain the complete boundary of the image. A brief introduction of ACO method is as follows.

### *3.1 Ant Colony Optimization*

ACO algorithm was introduced by Colorni et al. [26] in 1991 as computational intelligence technique to solve combinational optimization problem. This algorithm works based on the foraging behaviour of ants. The communications among population of

ants are indirect. The ants were communicated by depositing pheromone (chemical substance) on the ground and find the path between their nest and food source.

ACO algorithm has three phases such as initialization, pheromone update, and solution. After the completion of all the components, the global optimum solution is observed. Initially, all the ants haphazardly seek the best solution area and pheromone quantity is updated by using vicinity search of each ant. After the first iteration, the feasible solution space is minimized and it finds their routes in the limited space. The probability of searching the neighborhood ant is done by

$$b^k(i,\ \mathrm{j}) = \begin{cases} \dfrac{[s\ (i,\ j)]^{\alpha}.[t(i,\ j)]^{\beta}}{\sum\limits_{l \notin V_k} [s(i,\ l)]^{\alpha}.[t(i,\ l)]^{\beta}}, & \text{if} \quad j \notin v_k \\ 0, \text{otherwise} \end{cases} \tag{1}$$

where, $s(i,\ j)$ is the pheromone density, $\alpha$ and $\beta$ are relative information of the pheromone and $v_k$ is list of the neighbor node which is already visited by the ant. The pheromones are updated as follows.

$$s_{ij} = (1 - \rho)\, s_{ij} + \sum_{k} \Delta s_{ij}^{k} \tag{2}$$

where $s_{ij}$ is the amount of pheromone deposited, $\rho$ is the pheromone evaporation coefficient and $\Delta s_{ij}^{k}$ is the amount of pheromone deposited by the $k$th ant. ACO avoids premature convergence due to distributed computation. At each stage of iteration, it acquires suitable solution. It has collective interaction of population of agents. The step-by-step execution of ACO algorithm is exposed in Fig. 2.

The ACO algorithm suppresses the speckle noise and preserves the edge pixel from US image of prostate. It retains the edges with sensible amount of speckle-noise suppression which is more substantial for further analysis. BCRNN method is implemented to improve the boundary of an image and to acquire the Region of Interest (ROI) accurately. The implementation of BCRNN methodology is given in the Sect. 3.2.

## *3.2 BCRNN Method*

The proposed framework is illustrated in Fig. 1. To extract accurate boundary region from the prostate cancer image, BCRNN method is implemented. At first, ACO technique was applied to the enhance TRUS prostate image which is given as input to upcoming levels. The BCRNN serializes the given image into several different perspectives, then conducts shape prediction. All these shape predictions are then merged as a complete inference by a multi-view fusion strategy. Thus, the result

**ACO Algorithm for Prostate Ultrasound Medical Image Enhancement**

Input : Ultrasound Image with Speckle Noise

Output : Noise free Ultrasound Image

Step 1: Initialize the population of ants.

Step 2: Initialize the pheromone trail.

Step 3: Evaluate fitness of the each ant in the colony by using equation 3.

$$F(x) = \log(\log(E(I(x)))) * \frac{n\_edges(I(x))}{M * N} * H(I(x))$$

Step 4: Determine the best optimum solution among all ants in the colony.

Step 5: Update Pheromone trail by using formula

$$s_{ij} = (1-\rho)\, s_{ij} + \sum_{k} \Delta s_{ij}^{k}$$

Step 6: Determine the search direction for further search by the formula

$$b^{k}\ (i, \mathrm{j}) = \begin{cases} \dfrac{\left[s\ (i, j)\right]^{\alpha} \cdot \left[t(i, j)\right]^{\beta}}{\sum_{l \in V_k} \left[s(i, l)\right]^{\alpha} \cdot \left[t(i, l)\right]^{\beta}}, & if\ j \notin v_k \\ 0, & otherwise \end{cases}$$

Step 7: Construct new solution using obtained best optimum solution.

Step 8: Generate new colony. After each iteration the trails of pheromone evaporate

Step 9: Repeat steps 3 to step 8, until it satisfy the maximum number of Iterations

**Fig. 2** ACO algorithm for prostate image enhancement

is concatenated with original image and feed into the next level for more shape details. The shape prediction map is primarily set as an even distribution. And, the above process continues until convergence occurs on boundary prediction map. It is depicted in the following section.

### 3.2.1 BCRNN to Find Shape Inference

In the process of segmentation, to address the incompleteness of boundary in prostate image is most difficult task. The boundary completion of ultrasound image is reviewed in this section. Kimia et al. was implemented Geometric analysis of curvature to achieve the complete boundary detection of an image [27, 28] and Yosef et al. [29] proposed visual curve completion method to identify the missing parts of the boundary and to complete the gap in the boundary. Based on these methods, memory-guided inference procedure is proposed to complete the boundary of the image. So, it is identified that Recurrent Neural Network (RNN) is suitable for the above process and it was termed as Boundary Complete RNN (BCRNN).

### 3.2.2 Serialization and Deserialization Process

The ultrasound image is transformed into an interpretable sequence by using polar coordinate system around the image centre to generate a serialization process as

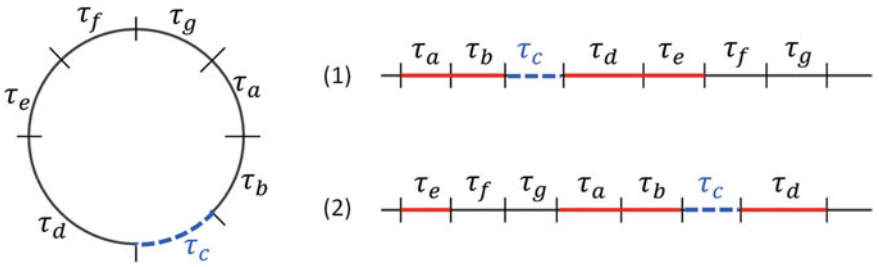



**Fig. 3** Serialization manner with different starting points

shown in Fig. 3. It is circle-wise manner of manual delineation. The serialized images are evenly partitioned into $T$ consecutive bands, which form the sequence as $x = (x_1, x_2, \ldots, x_T)$ and straightened version of $x$ is given as sequential input to the BCRNN process. Deserialization is the reverse process of serialization. In training process, all the segmentation label images are also serialized into sequence form.

### 3.2.3 Recurrent Neural Network

Given an input sequence of an image as $R$, hidden state vector $h = (h_1, h_2, \ldots, h_T)$ and output vector $S = (S_1, S_2, \ldots, S_T)$ and time step as $t = 1$ to $T$. The following equations are iterated for the time step

$$h_t = \xi(W_{Rh}R_t + W_h h_{t-1} + b_h) \tag{3}$$

$$S_t = W_{hS}h_t + b_y \tag{4}$$

where, $W$ denote recurrent neural network weight matrices, b denotes bias vector and $\xi$ is hidden layer function. Hidden state vector $h_t$ summarizes the information from previous $h_{t-1}$ and current input $R_t$, it can be exploited to infer current prediction $S_t$. In this process, the hidden state vector $h_t$ is considered as shape knowledge of the prostate image which is accrued from previous time steps. And, $h_t$ is used to deduce the boundary location for current time step t by taking an extra input $R_t$.

It is difficult to access long-range context in RNNs. Since the information stored in hidden layers take more time and consequently loses impact on future inference. Therefore, Long Short-Term Memory (LSTM) module is used to enhance hidden layer by controlling the flow of information in the network memory.

In estimating the missing parts of the boundary information are critical since it is from multiple directions. A single LSTM stream can make information from only one direction. For that reason, Bidirectional LSTM (BiLSTM) technique is implemented to control missing parts of the boundary and content of information of an image. Graves et al. [30] proposed BiLSTMmethodto acquire information from

both directions with two separate hidden layers and these are forwarded to the output layer. The computational of a BiLSTM at time step t are illustrated in the following equations (Eqs. 5–7).

$$\overrightarrow{h_t} = \xi\left(W_{x\overrightarrow{h}}R_t + W_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{h_{t-1}} + b_{\overrightarrow{h}}\right) \tag{5}$$

$$\overleftarrow{h_t} = \xi\left(W_{x\overleftarrow{h}}R_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h_{t-1}} + b_{\overleftarrow{h}}\right) \tag{6}$$

$$S_t = W_{\overrightarrow{h}S}\overrightarrow{h_t} + W_{\overleftarrow{h}S}\overleftarrow{h_t} + b_y \tag{7}$$

where, $\overrightarrow{h_t}$ is a forward hidden states and $\overleftarrow{h_t}$ is a backward hidden state by iterating the forward layer from $t = 1$ to $T$ and backward layer from $t = T$ to t respectively. The combinations of forward and backward hidden state are shown in Eq. 7. BiLSTM can combine with serialization and deserialization and formed BCRNN to yield accurate boundary trace from various directions. Thus, BCRNN can obtain incomplete boundary of prostate ultrasound images.

### 3.2.4 Multiple View Point Combination

While serializing the image from different starting points will create various shape predictions. Thus, the serializing from different starting points may change the relative distances between context-dependent sequence elements and it brings about slight difference in predictions. Figure 3 shown serialization with different starting points.

As shown in the above figure, consider $\tau_c$ is the missing boundary scrap then need indications from both $\tau_{a-b}$ and $\tau_{d-c}$ scrap to recuperate it. In the first step of serialization, it conserves the relative spatial relationship among three scrap. In the second step, destroys the continuity between $\tau_e$ and $\tau_d$ then $\tau_e$ has more distance from $\tau_c$. In the next step, BiLSTM needs to keep the $\tau_e$ information much longer time in the memory to achieve $\tau_c$. To overcome this problem, the original images are serialized from three view points and merge the complementary boundary predictions generated by BCRNN into a shape prediction. Thus, observing of an object in multiple viewpoints is noted as multiple view point fusion.

### 3.2.5 Refinement by Multiscale Auto-Context

It is further embeded BCRNN with Multiscale Auto-Context scheme [31] to gain consecutive refinement on the initial prediction by exploring the boundary information from neighbors. In BCRNN, level $k - 1$ was merged with prediction map

generated by original image and it is an input for level k. After the training of level $k-1$ and level $k$ is trained in BCRNN model. Standard classifiers embed into Auto-Context scheme to collect-related information [31, 32] and to control context information from near or far ranges.

Thus, BCRNN model has the inbuilt ability to flexibly leverage context information from near or far ranges. This capability benefits from memory are retained dynamically by BiLSTM. BCRNN along with Auto-Context scheme and multiscale mechanism are implemented to obtain informative maps with strong guidance for each level. For this process, iterative steps exploited to carry out multiscale Auto-Context scheme are shown in Eq. 8.

$$y^k = \kappa\left(\left(x^k; y^{k-1}\right), s^k\right) \tag{8}$$

where, $\kappa$ is model function of BCRNN, $y^k$ is the shape predication map from level $k$, $s^k$ is the scale used by BCRNN in level $k$ to generate the sequence of $x^k$. Finally, BCRNN along with multiscale Auto-Context scheme recovered the missing boundaries of the image. The step-by-step implementation of BCRNN algorithm is exposed in Fig. 4.

The experimental analysis of the proposed techniques is depicted in the following section.

**BCRNN Algorithm for Prostate Ultrasound Image Segmentation**

Input : De-speckle Ultrasound Image

Output : Segmented Image

Step 1: Convert image into polar coordinate

Step 2: Serialized images are partitioned into consecutive bands as

$x = (x_1, \ldots, x_T)$

Step 3: Segmented label images are also serialized into sequence form

Step 4: Recurrent Neural Network, the equations are iterated for the time step as $t = 1$ to $T$

$$h_t = \xi\left(W_{Rh} R_t + W_h h_{t-1} + b_h\right)$$

$$S_t = W_{hS} h_t + b_y$$

Step 5: In BiLSTM, at time step t the following equations are executed

$$\vec{h}_t = \xi\left(W_{x\vec{h}} R_t + W_{\vec{h}\vec{h}} \vec{h}_{t-1} + b_{\vec{h}}\right)$$

$$\overleftarrow{h}_t = \xi\left(W_{x\overleftarrow{h}} R_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}\right)$$

$$S_t = W_{\vec{h}S} \vec{h}_t + W_{\overleftarrow{h}S} \overleftarrow{h}_t + b_y$$

Step 6: Multi-view point combination are executed, as in the figure 3

Step 7: Multiscale Auto-context is implemented by the formula

$$y^k = \kappa\left(\left(x^k; y^{k-1}\right), s^k\right)$$

Step 9: Repeat steps 4 to step 7, until it satisfy the maximum number of Iterations

**Fig. 4** Implementation of BCRNN algorithm

## 4 Experimental Results

The various categories of TRUS prostate images are examined. The proposed method is evaluated and analyzed using parameters. The performance is measured by Dice Similarity coefficient, JI, Accuracy, Precision, Recall, and Boundary Error. The detailed explanation of experimental analysis and implementation are depicted in the following section.

### *4.1 Experimental Analysis*

For experimentation, 200 ultrasound prostate images with the size of 468 $\times$ 356 are used to assess the performance of TRUS image segmentation method. Initially, Ant Colony Optimization (ACO) method is used to remove the speckle noise and enhance the fine details in the US images. The preprocessed image is applied in BCRNN method to acquire exact shape of the image. The implementation details are depicted in the following section.

### *4.2 Implementation Details*

BCRNN has been trained with a many-to-many manner, in which direct mapping was constructed between input intensity sequence and the boundary label sequence. Long Term Short Memory (LSTM) was used and it consists of forward and backward streams with 500 hidden memory units. A Euclidean distance-based objective function has been used for training BCRNN and the network parameters are updated with Root Mean Square Propagation (RMSProp) [33] optimizer using the Back Propagation Through Time (BPTT) algorithm. The learning rate was set as 0.001 for all levels. All computational were conducted on a computer with Intel Core i5-5200U CPU@2.20 GHz processors.

## 5 Performance Evaluation

The results of the proposed method are shown in Fig. 5. The method obtains the shape of the prostate boundary without loss for some extends. It also conquers the best shape of the boundary than the manual segmentation.

**Quantitative Analysis**

Performance of the proposed work ais evaluated by evaluation matrices such as Dice Similarity Coefficient (Dice), Jaccard Index (JI), Precision, Recall, Accuracy,

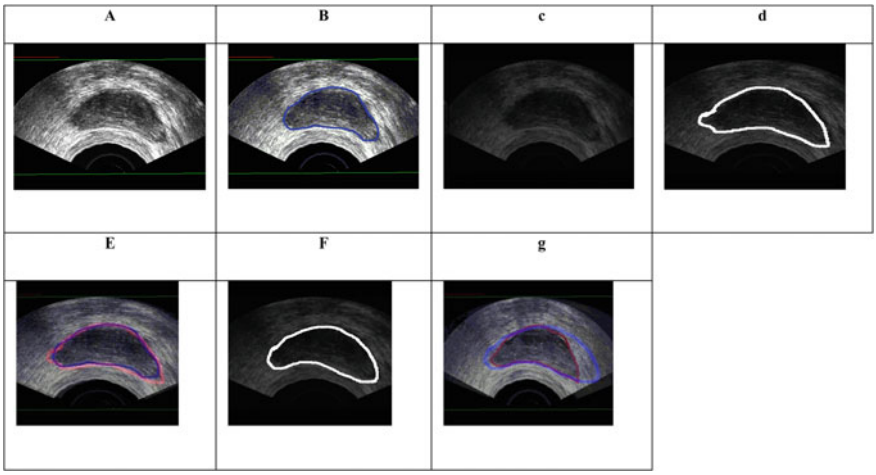



**Fig. 5** Qualitative result of the various methods. **a** Original image, **b** ground truth image, **c** enhanced Image using ACO, **d** BCRNN, **e** BCRNN with ground truth and **f** ACO_BCRNN, **g** ACO_BCRNN with ground truth

and Boundary Error. The quantitative results produced by the method BCRNN and BCRNN with ACO for various numbers of TRUS images are tabulated in Table 1.

The average of various performance measures of the proposed segmentation model is exposed in Table 2. From the Table 2, it is observed that the Dice Similarity Coefficient (Dice) of the BCRNN_ACO is 0.9801, which is 0.11% higher than BCRNN. The pictorial representation of Dice similarity earned by BCRNN and ACO with BCRNN is shown in Fig. 6.

The BCRNN with ACO showed superior performance in terms of accuracy, since the accuracy of ACO _BCRNN method is 0.1622% higher than accuracy earned by BCRNN method. The comparison chart is revealed in Fig. 7.

The performance of ACO with BCRNN is also validated by Jaccard Index (JI). BCRNN with ACO yielded JI of 0.9996, which is 0.0002 is lesser than the BCRNN method. Its graphical representation is shown in Fig. 8.

The precision of BCRNN is 95.95%, which is 0.2% lower than ACO with BCRNN approach. The pictorial representation of precision measure is shown in Fig. 9.

The Boundary Error (BE) of proposed method is 4.0841while BCRNN method earned 4.3255. It is observed that ACO with BCRNN method achieved less error, which is 0.2414 lesser than ACO. It proves that the proposed method achieved more boundary pixels. The comparison chart of the Boundary Error is shown in Fig. 10. The recall of the proposed method and BCRNN is exposed in Fig. 11. It clearly proved that both the methods yielded similar performance. The recall of ACO and ACO with BCRNN is 99.96%. From the above experimental analysis, it is absorbed that the proposed ACO_BCRNN method yields a better result that proves for better segmentation.

**Table 1** Quantitative results of segmentation methods

| | Evaluation metrics | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of images | Dice | | JI | | Acc | | Precision | | RECALL | | BE | |
| | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO |
| 10 | 0.9822 | 0.9808 | 1.0000 | 1.0000 | 96.6168 | 96.3761 | 0.9654 | 0.9628 | 0.9996 | 0.9996 | 3.6340 | 3.9162 |
| 20 | 0.9827 | 0.9798 | 1.0000 | 1.0000 | 96.7080 | 96.1903 | 0.9664 | 0.9608 | 0.9996 | 0.9996 | 3.5283 | 4.1335 |
| 30 | 0.9828 | 0.9811 | 1.0000 | 1.0000 | 96.7324 | 96.4236 | 0.9666 | 0.9633 | 0.9995 | 0.9995 | 3.4999 | 3.8610 |
| 40 | 0.9827 | 0.9806 | 1.0000 | 1.0000 | 96.7080 | 96.3486 | 0.9664 | 0.9625 | 0.9995 | 0.9995 | 3.5282 | 3.9485 |
| 50 | 0.9829 | 0.9810 | 1.0000 | 1.0000 | 96.7532 | 96.4078 | 0.9668 | 0.9631 | 0.9995 | 0.9995 | 3.4757 | 3.8794 |
| 60 | 0.9835 | 0.9819 | 1.0000 | 1.0000 | 96.8617 | 96.5703 | 0.9680 | 0.9649 | 0.9995 | 0.9995 | 3.3527 | 3.6935 |
| 70 | 0.9840 | 0.9825 | 1.0000 | 1.0000 | 96.9425 | 96.6863 | 0.9689 | 0.9661 | 0.9995 | 0.9995 | 3.2612 | 3.5606 |
| 80 | 0.9843 | 0.9830 | 1.0000 | 1.0000 | 96.9995 | 96.7760 | 0.9695 | 0.9671 | 0.9995 | 0.9995 | 3.1965 | 3.4580 |
| 90 | 0.9845 | 0.9834 | 1.0000 | 1.0000 | 97.0482 | 96.8431 | 0.9700 | 0.9678 | 0.9995 | 0.9995 | 3.1414 | 3.3811 |
| 100 | 0.9847 | 0.9837 | 1.0000 | 1.0000 | 97.0838 | 96.8962 | 0.9704 | 0.9684 | 0.9995 | 0.9995 | 3.1010 | 3.3202 |
| 110 | 0.9848 | 0.9838 | 1.0000 | 1.0000 | 97.0923 | 96.9203 | 0.9705 | 0.9686 | 0.9995 | 0.9995 | 3.0911 | 3.2920 |
| 120 | 0.9848 | 0.9839 | 1.0000 | 1.0000 | 97.0927 | 96.9338 | 0.9705 | 0.9688 | 0.9995 | 0.9995 | 3.0902 | 3.2759 |
| 130 | 0.9846 | 0.9838 | 1.0000 | 1.0000 | 97.0593 | 96.9116 | 0.9701 | 0.9685 | 0.9995 | 0.9995 | 3.1289 | 3.3014 |
| 140 | 0.9847 | 0.9839 | 1.0000 | 1.0000 | 97.0706 | 96.9326 | 0.9703 | 0.9688 | 0.9995 | 0.9995 | 3.1158 | 3.2771 |
| 150 | 0.9847 | 0.9839 | 1.0000 | 1.0000 | 97.0735 | 96.9439 | 0.9703 | 0.9689 | 0.9995 | 0.9995 | 3.1123 | 3.2637 |
| 160 | 0.9840 | 0.9832 | 1.0000 | 1.0000 | 96.9562 | 96.8153 | 0.9690 | 0.9674 | 0.9995 | 0.9995 | 3.2510 | 3.4229 |
| 170 | 0.9831 | 0.9820 | 1.0000 | 1.0000 | 96.7992 | 96.6193 | 0.9672 | 0.9652 | 0.9996 | 0.9996 | 3.4481 | 3.6724 |
| 180 | 0.9812 | 0.9808 | 0.9999 | 0.9998 | 96.4973 | 96.4121 | 0.9638 | 0.9629 | 0.9996 | 0.9996 | 3.8453 | 3.9343 |
| 190 | 0.9795 | 0.9801 | 0.9998 | 0.9997 | 96.2058 | 96.2983 | 0.9604 | 0.9616 | 0.9996 | 0.9996 | 4.2291 | 4.0763 |

(continued)

**Table 1** (continued)

| | Evaluation metrics | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of images | Dice | | JI | | Acc | | Precision | | RECALL | | BE | |
| | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO | BCRNN | BCRNN with ACO |
| 200 | 0.9790 | 0.9801 | 0.9998 | 0.9996 | 96.1266 | 96.2888 | 0.9596 | 0.9615 | 0.9996 | 0.9996 | 4.3255 | 4.0841 |

**Table 2** Relative quantitative results of segmentation method

| Methods | Dice | JI | ACC | Precision | Recall | BE |
|---|---|---|---|---|---|---|
| BCRNN | 0.9790 | 0.9998 | 96.1266 | 0.9596 | 0.9996 | 4.3255 |
| **ACO_BCRNN** | **0.9801** | **0.9996** | **96.2888** | **0.9615** | **0.9996** | **4.0841** |

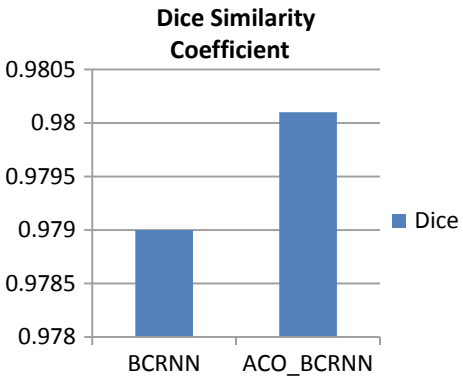



**Fig. 6** Dice similarity co-efficient result

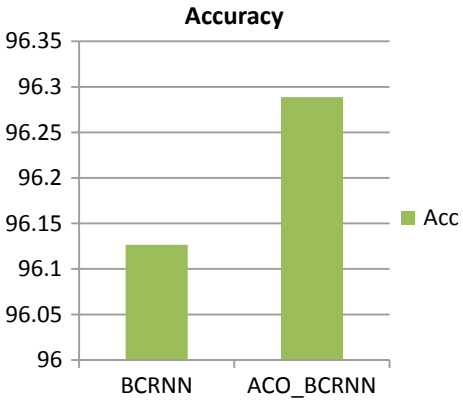



**Fig. 7** Accuracy result

## 6 Conclusion

Segmentation framework for prostate ultrasound images is proposed in this paper. The TRUS image was first enhanced with Ant Colony Optimization (ACO) method to remove the speckle noise. Subsequently, the boundary of the image was extracted without loss by using Boundary Complete Recurrent Neural Network (BCRNN). The expected boundary and region of interest of the prostate in TRUS image are obtained using the proposed method. The performance of the segmented framework is evaluated by evaluation metrics. The experimental result clearly showed that the performance of the proposed segmented method ACO with BCRNN is superior

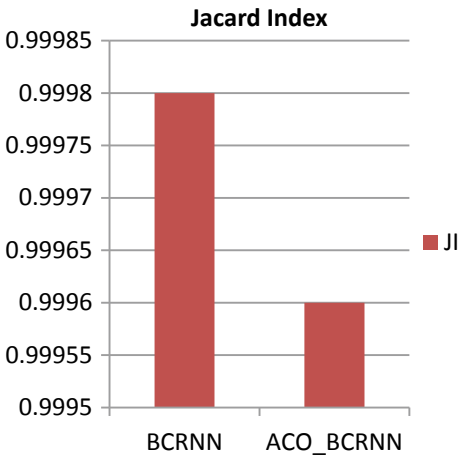



**Fig. 8** Result of Jacard index

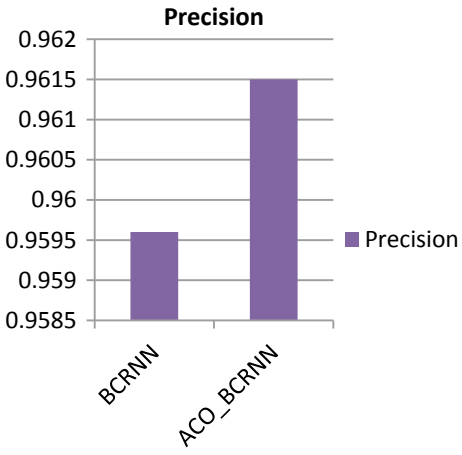



**Fig. 9** Result of precision

compared to BCRNN in terms of Dice Similarity Coefficient, Jacard Index, Accuracy, Precision, Recall, and Boundary Error. Thus the quantitative result also proves that ACO_BCRNN is an outstanding tool for prostate image segmentation.

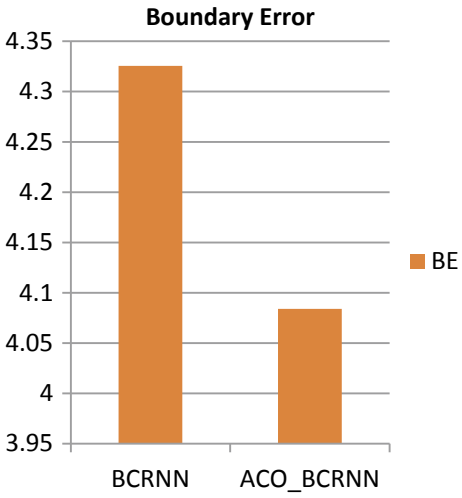



**Fig. 10** Performance of boundary error

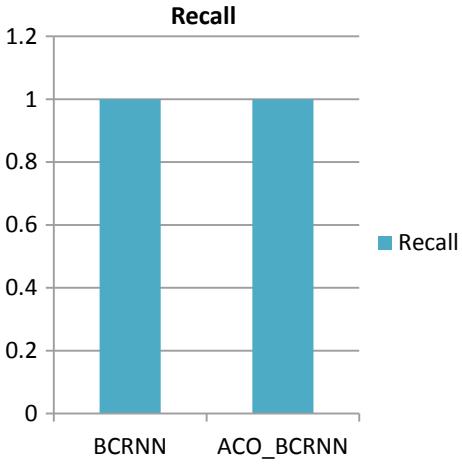



**Fig. 11** Result of recall

## References

1. Cancer facts and figures. American Cancer Society [Online]. https://www.cancer.org (2020)
2. Aldasoro CCR, Abhir B (2007) Volumetric texture segmentation by discriminant feature selection andmultiresolution classification. IEEE Trans Med Imaging 26:1–14
3. Olivier J, Cedric M, Jean JR, Romuald B, Hubert C (2008) A supervised texture-based active contour model with linear programming. In: Proceedings of 15th IEEE international conference image processing, pp 1104–1107
4. Schaap M, Theo W, Lisan N, Coert M, Ermanno C, de Marleen B, Wiro N (2011) Robust shape regression for supervised vessel segmentation and its application to coronary segmentation in CTA. IEEE Trans Med Imaging 30:1974–1986

5. Lee NR, Theodore S, Andrew FL (2008) Interactive segmentation for geographic a trophy in retinal fundus images. In: Conference record/Asilomar conference signals, systems and computers, pp 655
6. Grau V, Mewes AUJ, Alcaniz M, Ron K, Simon KW (2004) Improved watershed transform for medical image segmentation using prior information. IEEE Trans Med Imaging 23:447–458
7. Fan S, Keck LV, Wan NS, Ruo WY (2003) Prostate boundary detection from ultrasonographic images, pp 605–623
8. Wu RY, Keck L, Wan NS (2000) Automatic prostate boundary recognition in son graphic images using feature model and genetic algorithm. J Ultrasound Med 19:771–782
9. Chiu B, Fenster A (2004) Prostate segmentation algorithm using dyadic wavelet transform and discrete dynamic contour. Phys Med Biol 49:4943–4960
10. Cruz-Aceves I, Aviña-Cervantes JG, López-Hernández JM, González-Reyna SE (2013) Multiple active contours driven by particle swarm optimization for cardiac medical image segmentation. Comput Math Methods Med
11. Otsu N (1979) A threshold selection method for grey level histograms. IEEE Trans Syst ManCybernet SMC 9(1):62–66
12. Kapur JN, Sahoo PK, Wong AKC (1985) A new method for gray-level picture thresholdingusing the entropy of the histogram. Comput Vis Graph Image Process 29:273–285
13. Luca AD, Termini S (1972) Definition of a non probabilistic entropy in the setting of fuzzy sets theory. Inf Contr 20:301–315
14. Bloch I (2005) Fuzzy spatial relationships for image processing and interpretation: a review. Image Vis Comput 3(2):89–110
15. Zhao MS, Fu AMN, Yan H (2001) A technique of three level thresholding based on probability partition and fuzzy 3-partition. IEEE Trans Fuzzy Syst 9(3):469–479
16. Tao WB, Tian JW, Liu J (2003) Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm. Pattern Recogn Lett 24:3069–3078
17. Cao L, Bao P, Shi Z (2008) The strongest schema learning GA and its application to multi-level thresholding. Image Vis Comput 26:716–724
18. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models-their training and application. Comput Vis Image Underst 61(1):38–59
19. Shen D, Zhan Y, Davatzikos C (2003) Segmentation of prostate boundaries from ultrasound images using statistical shape model. IEEE Trans Med Imaging 22(4):539–551
20. Van Ginneken B, Frangi AF, Staal JJ, Ter Haar Romeny BM, Viergever MA (2002) Active shape model segmentation with optimal features. IEEE Trans Med Imaging 21(8):924–933
21. Rogers M, Graham J (2002) Robust active shape model search. In European conference on computer vision. Springer, pp 517–530
22. Santiago C, Nascimento JC, Marques JS (2015) 2D segmentation using a robust active shape model with theem algorithm. IEEE Trans Image Process 24(8):2592–2601
23. Zhou X, Huang X, Duncan JS, Yu W (2013) Active contours with group similarity. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2969–2976
24. Yan P, Xu S, Turkbey B, Kruecker J (2010) Discrete deformable model guided by partial activeshape model for trus image segmentation. IEEE Trans Biomed Eng 57(5):1158
25. Yang X, Yu L, Yu L, Wang Y, Ni D, Qin J, Heng P-A (2017) Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: Proceedings of the thirty-first AAAI conference on artificial intelligence
26. Colorni A, Dorigo M, Maniezzo V (1991) Distributed optimization by ant colonies. In: Proceedings of ECAL'91 European conference on artificial life. Elsevier Publishing, Amsterdam, The Netherlands, pp 134–142
27. Kimia BB, Frankel I, Popescu AM (2003) Euler spiral for shape completion. Int J Comput Vis 54(1–3):159–182
28. Rueda S, Knight CL, Papageorghiou AT, Noble JA (2015) Feature-based fuzzy connectedness segmentation of ultrasound images with an object completion step. Med Image Anal 26(1):30–46

29. Ben-Yosef G, BenShahar O (2012) A tangent bundle theory for visual curve completion. IEEE Trans Pattern Anal Mach Intell 34(7):1263–1280
30. Graves A, Jaitly N, Mohamed A (2013) Hybrid speech recognition with deep bidirectional lstm. In: IEEE workshop on automatic speech recognition and understanding (ASRU), pp 273–278
31. Tu Z, Bai X (2010) Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans Pattern Anal Mach Intell 32(10):1744–1757
32. Gao Y, Wang L, Shao Y, Shen D (2014) Learning distance transform for boundary detection and deformable segmentation in CT prostate images. In: International workshop on machine learning in medical imaging. Springer, pp 93–100
33. Tieleman T, Hinton G (2012) Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSER A Neural Netw Mach Learn 4(2)

# A Deep Learning Approach to Detect Lumpy Skin Disease in Cows


**Gaurav Rai, Naveen, Aquib Hussain, Amit Kumar, Akbar Ansari, and Namit Khanduja**

**Abstract** Diseases in cows are an influential point for human concern. There are some diseases in animals identified in the early phases that can be diagnosed and cured in the early phases of the disease itself. The effect of lumpy skin disease can cause large capital losses in the farm animal industry if it is not taken care of properly. The main reason for this disease is the lumpy skin virus, and this virus is a part of the Poxviridae family. The major symptom of lumpy skin disease is the Neethling strain, and other symptoms are a few mild forms of circumscribed skin nodules. These symptoms also include mucous membranes of internal organs like respiratory organs and reproductive organs. By the infection of such disease, animals like cattle get their skin permanently damaged. Some of the detrimental outcomes of this disease in cows are reduction in milk projection, infertility, poor growth, abortion and sometimes death. In this research work, an architecture using machine learning techniques to detect the disease is proposed. This architecture employs the pre-trained models like VGG-16, VGG-19 and Inception-v3 for feature extraction and then followed by multiple classifiers. The work is tested on our manually collected dataset, and the extracted features were further classified using the classifiers like kNN, SVM, NB, ANN and LR. Using this methodology, the state-of-the-art solution obtaining a classification accuracy of 92.5% over the test dataset.

**Keywords** Feature extraction · Dataset collection · Deep learning · Lumpy skin disease · Neural network (NN)

## 1 Introduction

Skin is an important part of an animal body. Lumpy skin disease (LSD) is a vigorous disease in cows extended by biting insects. LSD is an acute noxious defect of cows

G. Rai (✉) · Naveen · A. Hussain · A. Kumar · A. Ansari · N. Khanduja
GKV, Haridwar, Uttarakhand, India

GEU, Dehradun, Uttarakhand, India

COER, Roorkee, Uttarakhand, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_30

of all ages. The LSD disease is identified by large skin knots covering all parts of the body leading to fever, nasal discharge, spread lymph nodes and lachrymation. Lumpy skin disease is generally found in Africa, Russia, Egypt, Oman and India. At first, it was identified in Egypt. This disease was earlier contained in Africa and went outside it in 1984. This disease has now moved to Madagascar and now in countries like the Arab Gulf Peninsula and the Middle East. The countries (Jordan, Syria, Lebanon, Turkey, Iran, and Iraq), which earlier did not have any cases of this disease, are now also suffering economic losses [1]. Soon after this, the lumpy skin starts to develop over the whole body or sometimes only develops around the neck, limbs and genitals region. These lesions are first seen around circumscribed areas around erect hairs. Around the normal skin, these lesions are found. The narrow rings formed around these lesions are haemorrhage. These lesions cover epidermis and adjacent subcutis [2]. The main reason for the infection is lesions on skin, but viruses can also spread through various secretions from the body like excretions and semen. Thereby, the potential targets of the virus are majorly targeted by physical means from hematophagous arthropods, and this involves bites from mosquitoes and flies and ticks [1, 2]. Transcendence and transmission capabilities of types of ticks are also possible. Characteristic disease lesions erupt after the incubation period of a couple of weeks after the infection under the testing circumstances, while it takes up to 2–5 weeks in ordinary cases [3]. Unluckily, there are no particular antiviral drugs convenient for the rectification of LSD. The auxiliary care is the only medication available for cows. It is essential to think about the rectification of skin lesions using wound care, sprays and the use of antibiotics to avoid secondary skin infections and pneumonia [4]. There are various parts of the paper in which Sect. 1 is introduction, Sect. 2 is data collection, Sect. 3 is the proposed approach, Sect. 4 is result, and Sect. 5 is conclusion [5]. Cow images infected from lumpy disease and normal cow image (uninfected) are given in Fig. 1.


**Fig. 1** Images of lumpy skin and normal skin in cows

**Table 1** Image dataset of cows

| Types of cow | Number of images |
|---|---|
| Infected from lumpy disease | 132 |
| Normal images of cow | 199 |

## 2 Data Collection

The collection of dataset is an important phase because there is no dataset available of lumpy skin disease in cows. So, we created an image dataset of cows around 300–400 images that is shown in Table 1.

Since the problem with the dataset is that it can be biased due to lack of proper dimensions and resolution; but in this dataset, we have chosen only high resolution and proper pixel images [6].

## 3 Proposed Approach

In this framework, we used deep convolutional neural networks. For extracting features from the images, we have used the pre-trained models like VGG-16, VGG-19 and Inception-V3. VGG-16 is a model composed of 16-convolutional layer, 5 max pooling layers. After this, we have applied all three properly associated layers, and RELU activation is then projected to all hidden layers. VGG-19 contains 19 layers which comprise 16 convolution layers and three fully connected layer, one softmax layer and five MaxPool layers. Inception-V3 is a CNN model which has 48 layers. The pre-trained model Inception-V3 is trained on more than 1 Million images of ImageNet database [4]. It can classify more than thousand object classes. We take an image as input, process it and classify it in various categories [6]. In this process, it consists of many layers in which the first layer is a convolution layer that operates the feature extraction by using filter or kernel. Since the images are nonlinear, so to remove the linearity of images, we used rectifier linear unit (ReLU) function [7]. Here, patches created for convolution operation and then converted in dimensional array. In this convolution process, all the layers are connected, and it has 1D inputs. Continue to the convolution process, linear transformation takes place, and the following derivatives are backward propagation process applied fully connected convolution layer [5].

$$E = \left(y' - O\right)^2/2$$
$$\frac{\partial \boldsymbol{E}}{\partial \boldsymbol{O}} = -\left(y' - O\right)$$
$$\frac{\partial \boldsymbol{O}}{\partial \boldsymbol{Z}2} = (O)(1 - O)$$

$$\frac{\partial \boldsymbol{Z}2}{\partial \boldsymbol{O}} = A1$$

Chain rule obtains the overall modifications in error with respect to the weight. The weight matrix depends on the gradient value.

$$W\_\text{new} = W\_\text{old} - lr * \frac{\partial E}{\partial W}$$

Weight matrix covers the backward propagation and that applied in connected layers and updated filters as follows:

$$\begin{aligned}
\frac{\partial E}{\partial f} &= \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial Z2} \cdot \frac{\partial Z2}{\partial A1} \cdot \frac{\partial A1}{\partial Z1} \cdot \frac{\partial Z1}{\partial f} \\
\frac{\partial E}{\partial O} &= -\left(y' - O\right) \\
\frac{\partial O}{\partial Z2} &= (O)(1 - O) \\
\frac{\partial Z2}{\partial A1} &= W^T \\
\frac{\partial A1}{\partial Z1} &= A1(1 - A1) \\
\frac{\partial Z1}{\partial f} &= X
\end{aligned}$$

The step by step working of CNN is shown in Fig. 2.

The weights, which are negative, show inhibitions in connection; meanwhile, the weights, which are positive, show excitation in connections. The following activity of the neuron cell is shown by our model [8]. All the inputs are added and are then updated by weights. This is linear combination. The amplitude of our function is then calculated by our activation function. For example, the most generic output ranges between 0–1 and −1 to 1. This step is described mathematically here in Fig. 3.

From this model, the interval activity of the neuron can be shown to be:

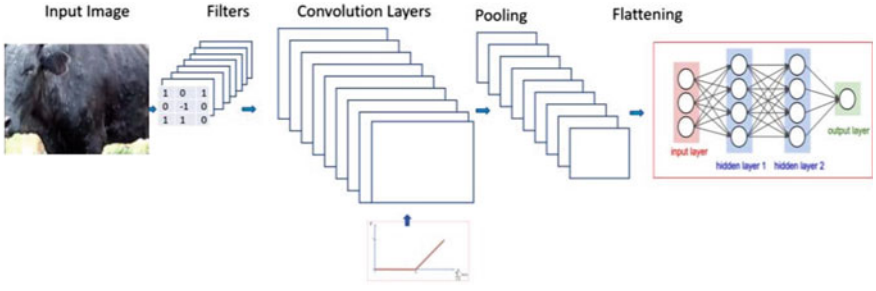



**Fig. 2** Working of convolutional neural network

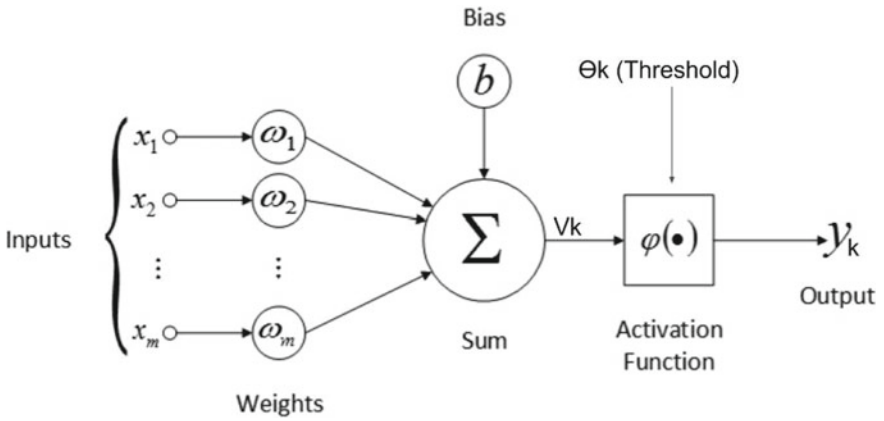



**Fig. 3** Working of artificial neural network (ANN)

$$v_k = \sum_{j=1}^{p} w_{kj} x_i$$

**Activation functions**

The yield of the neuron ($y_k$) would be the same as the output on the data input of $v_k$ on the activation function. As earlier studied, activation function works the same way as a compressing function, that is the yield of the neuron in the NN among assured values generally 0 and 1, or −1 and 1. Commonly, we have three ways in which activation function works, represented by Φ(.). The first function is known as threshold function which gives output 0 if the input is less than zero and gives output 1 if the input is greater than or equal to zero [9].

$$\phi(v) = 1 \quad \text{if} \quad v \geq 0$$
$$\phi(v) = 0 \quad \text{if} \quad v < 0$$

Secondly, the linear function takes in input in the range of 0 or 1, but sometimes, it takes values dependent upon the amplification factor inside the region of linear operation.

$$\phi(v) = 1 \quad \text{if} \quad v \geq \frac{1}{2}$$
$$\phi(v) = v \quad \text{if} \quad -\frac{1}{2} > v > \frac{1}{2}$$
$$\phi(v) = 0 \quad \text{if} \quad v < -\frac{1}{2}$$

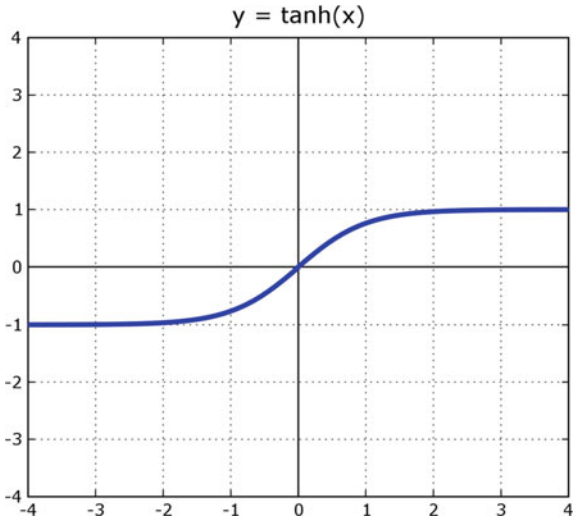



**Fig. 4** Graph of hyperbolic tangent function

Thirdly, the sigmoid function can vary between 0 and 1, and sometimes between −1 and 1 range is more useful.

$$\phi(v) = \tanh\left(\frac{v}{2}\right)$$
$$\phi(v) = \frac{1 - \exp(-v)}{1 + \exp(-v)}$$

The ANN represented by us consists of all variations on the parallel assigned processing idea. The design of every neural network established on much alike building blocks which accomplish the processing. In next, we will discuss these processing units and various neural network topologies (Fig. 4).

Image embedder uses different models such as VGG-16, VGG-19 and Inception-V3 for feature extraction. We split the dataset into 75% as training data and rest 25% as testing data [6].

We evaluate the ROC curve for the calculation of performance on our framework. On *X*-axis "Specificity" is defined, and on *Y*-Axis, "Sensitivity" is mentioned. Figure 5 represents the ROC curve on lumpy skin disease, and Fig. 6 represents the ROC curve on normal cow images [9].

## 4 Results

In this section, we come up with the final result achieved by the tests done in the above section. In this project, we have used three scenarios. In the first scenario, Inception-v3 is used for feature extraction, and then with the help of various classifiers like

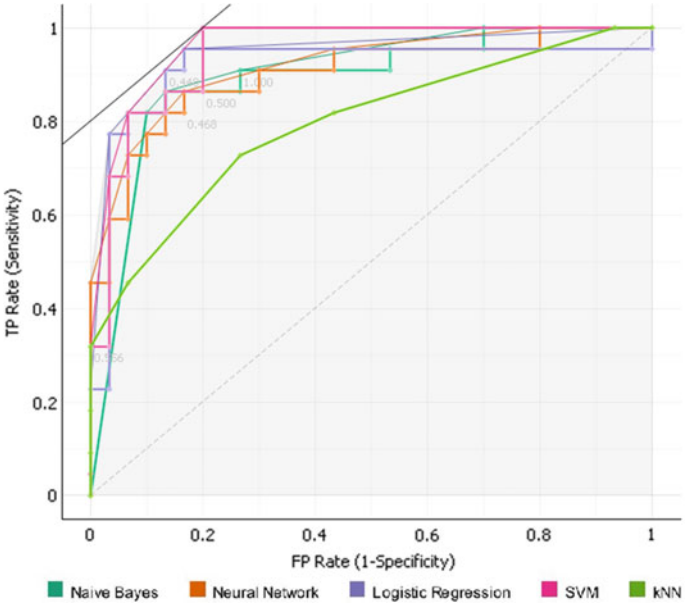



**Fig. 5** ROC curve on lumpy skin disease

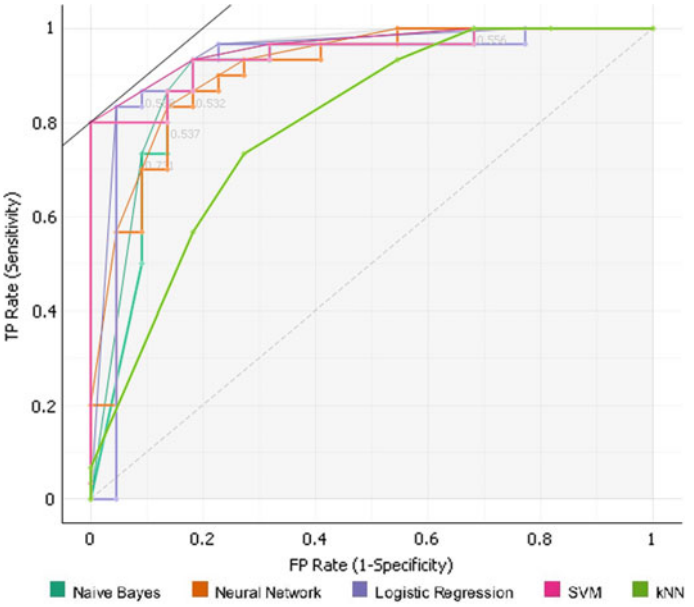



**Fig. 6** ROC curve on normal cow images

**Table 2** Accuracy assessment

| Feature extractor | Model | Accuracy | AUC | Precision |
|---|---|---|---|---|
| Inception-v3 | SVM | 0.830 | 0.876 | 0.830 |
| | KNN | 0.698 | 0.848 | 0.738 |
| | Neural network | 0.925 | 0.955 | 0.926 |
| | Naïve Bayes | 0.717 | 0.783 | 0.720 |
| | Logistic regression | 0.887 | 0.947 | 0.887 |
| VGG-16 | SVM | 0.848 | 0.889 | 0.857 |
| | KNN | 0.848 | 0.906 | 0.854 |
| | Neural network | 0.864 | 0.859 | 0.862 |
| | Naïve Bayes | 0.833 | 0.863 | 0.831 |
| | Logistic regression | 0.879 | 0.830 | 0.877 |
| VGG-19 | SVM | 0.765 | 0.842 | 0.807 |
| | KNN | 0.765 | 0.882 | 0.780 |
| | Neural network | 0.838 | 0.870 | 0.838 |
| | Naïve Bayes | 0.882 | 0.867 | 0.882 |
| | Logistic regression | 0.853 | 0.876 | 0.853 |

KNN, SVM, ANN, LR and NB, artificial neural network gives the maximum possible classification accuracy of 92.5% as shown in Table 2. In the second scenario, we have used VGG-16 for feature extraction and achieved a maximum possible accuracy of 87.9% with the help of logistic regression among all classifiers used. In the third scenario, a 19-layer convolutional neural network (VGG-19) is used for extracting the features and achieved the maximum possible accuracy of 88.2% with the help of Naive Bayes among all classifiers used. By observing all the three scenarios, we perceived the best combination is of Inception-v3 with artificial neural network giving us the highest possible classification accuracy of 92.5% over the test dataset. The accuracy assessment table is given (Table 2).

## 5 Conclusion

Deep convolutional neural network is proposed in this research work to predict lumpy skin and normal skin in cows. The results demonstrate that for a given dataset, the proposed model attains an accuracy of 92.5%. Since there is no standard dataset available, there are many problems that can arise; but in this research work, standard quality of the image dataset is used, due to which the proposed model attains better performances. This model can be used in different fields of medical science like to detect skin cancer in animals and other diseases. Moreover, it can help veterinary surgeons to figure out animal disease problems in early stages that do not require much manual work.

**Acknowledgements** We would like to express our special thanks to Dr. Ankush Mittal (300+ publications and 2 international Books) and Mr. Namit Khanduja for the proper guidance and motivation at various stages.

## References

1. Annandale C, Holm D, Ebersohn K, Venter E (2013) Seminal transmission of lumpy skin disease virus in heifers. Trans bound Emerg Dis 61. https://doi.org/10.1111/tbed.12045
2. Tuppurainen E (2017) Epidemiology of lumpy skin disease
3. Shadab M, Dwivedi M, Omkar SN, Javed T, Bakey A, Raqib M, Chakravarthy A (2019) Disease recognition in sugarcane crop using deep learning. https://doi.org/10.13140/RG.2.2.21849.47209
4. Aishwarya AG, Nijhawan R (2019) A deep learning approach for classification of onychomycosis nail disease. https://doi.org/10.1007/978-3-030-30577-2_98
5. Varshni D, Thakral K, Agarwal L, Nijhawan R, Mittal A (2019) Pneumonia detection using CNN based feature extraction, pp 1–7. https://doi.org/10.1109/ICECCT.2019.8869364
6. Chhabra H, Srivastava A, Nijhawan R (2019) A hybrid deep learning approach for automatic fish classification. https://doi.org/10.1007/978-3-030-30577-2_37
7. Nijhawan R, Joshi D, Narang N, Mittal A, Mittal A (2019) A futuristic deep learning framework approach for land use-land cover classification using remote sensing imagery. https://doi.org/10.1007/978-981-13-0680-8_9
8. Nijhawan R, Rishi M, Tiwari A, Dua R (2019) A novel deep learning framework approach for natural calamities detection: proceedings of third international conference on ICTCS 2017. https://doi.org/10.1007/978-981-13-0586-3_55
9. Kozma R, Ilin R, Siegelmann H (2018) Evolution of abstraction across layers in deep learning neural networks. Proc Comput Sci 144:203–213. https://doi.org/10.1016/j.procs.2018.10.520

# Prediction of Influenza-like Illness from Twitter Data and Its Comparison with Integrated Disease Surveillance Program Data


**Monica Malik and Sameena Naaz**

**Abstract** The social networking sites are currently assisting in delivering faster communication and they are also very useful to know about the different people's opinions, views, and their sentiments. Twitter is one of the social networking sites, which can help to predict many health-related problems. In this work, sentiment analysis has been performed on tweets to predict the possible number of cases with H1N1 disease. The data will be collected country wise, where the tweets lie between four ranges on which the further analysis will be done. The results show the position of India based on the frequency of occurrence in the tweets as compared to the other countries. This type of disease prediction can help to take a quick decision in order to overcome the damage. The results predicted by sentiment analysis of Twitter data will then compared with the data obtained from the 'Ministry of Health and Family Welfare-Government of India' site. The data present at this site gives the actual number of cases occurred and collected by Indian Governments "Integrated Disease Surveillence Program". Comparison with this data will help in calculating the accuracy of the sentiment analysis approach proposed in this work.

**Keywords** Influenza · Swine flu · H1N1 · Decongestants · Rapivab · Tamiflu · Relenza · Flomist · Zanamivir

## 1 Introduction

The internet is one of the vital assets that can help in the prediction of illness outbursts. It provides a chance for low-cost time-sensitive bases to be demoralized in order to increase the existing surveillance systems. In India, surveys are most prevalent strategies in various fields to gauge the public sentiments with respect to diseases. Survey-based systems are expensive and tedious process and are not reasonable for a transferable pandemic. One of the conceivable answers to beat this issue is

M. Malik · S. Naaz (✉)
Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India
e-mail: snaaz@jamiahamdard.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_31

utilization of web-based social networking sites which permits quick trade of ideas, thoughts, and data utilizing web-based applications. Online networking has been an essential focal point of the data recovery and content mining field from the beginning, since it produces enormous unstructured textual data and presents users relations progressively. There are a rising number of online networking tools and a quickly developing client base over all socioeconomics. Social networking sites have turned out to be quick and less expensive that empowers brisk and simple access to open data among potential clients.

There are many social networking sites which are being used worldwide and is huge a platform for everyone to communicate like Facebook, Quora, Twitter, and Instagram (where the emotions and views are expressed via photos and captions). Talking about Twitter, it is one of the leading social networking sites of twenty-first century. It's a free of cost networking plus micro-blogging platform that allows its huge number of users to share their views and exchange views and tweets. Twitter has around 330 million active users every month where 500 million tweets are tweeted per day and around 200 billion tweets per year are encountered. Useful information about events happening around the world, the news, and everything is posted by the users on this micro-blogging site. Social media is producing huge amount of data. People share their views, experiences, and opinions on many different topics. Users share their reviews on the products that they have ordered or they have used, they share their views on current affairs, if a group wants to advertise their events and/or campaigns then that is done by socializing on Twitter via tweets, the users tell their opinions about the movie they have watched, they share their experiences like the customers who have faced problems using a particular product and the problem wasn't resolved on multiple phonic conversations then they can post their complaint on Twitter by mentioning their problem and the company's name on a tweet.

Individuals look for wellbeing data for any number of reasons: worry about themselves [1], their family [2] or their companions [3]. Some quests are just because of general intrigue, maybe affected by a news report or a current logical production, and postponements in identifying the beginning of an irresistible plague brings about a major harm towards a society. Therefore, there is solid enthusiasm for decreasing these postponements. One approach to achieve this is through a surveillance framework, which emphasizes the use of the real-time data analysis [4] which recognize and describes unfamiliar action for further public health issues [5, 6] With the increase in the usage of online networking sites [7, 8], for example, Twitter, screening of the pandemic H1N1 infection has been done in [9].

In their work [10] proposed a novel procedure considering dynamic watchwords from RSS channels which are utilized to recover tweets. Conclusion examination and tally construct system relate to informational indexes to look at essential issues identified with Influenza-A (H1N1) pandemic. The approach used in their paper focuses on various parameters to locate the important data towards a specific sickness and the general mindfulness towards it. Twitter information has been utilized to quantify political assessment in [11], used to forecast stock returns [12], national assumption (i.e., joy) in [13], people's sentiments about outdoor game 'Lawn Tennis' in [14] and to screen the effect of seismic tremor in [4]. Comparison of traditional epidemiology has been made with computational epidemiology in [15].

The organization of rest of the paper is as follows: Sect. 2 discusses some of the similar work reported in the literature. The approach proposed in this work is discussed in Sect. 3. Section 4 gives the results and discussions which is followed by conclusion in Sect. 5. At the end limitations of the work and future scope is discussed in Sect. 6.

## 2 Related Work

Twitter is a free informal communication and a blogging platform that empowers its huge number of clients to exchange each other's "tweets" constrained to 140 characters. Clients decide if their tweets can be viewed by the overall population or are limited to some predefined viewers only. The platform has more than 190 million enrolled clients and produces around 55 million tweets every day. A current investigation of the "Twitter stream" uncovered that a generous extent of tweets contains general discussion about various aspects of life. Despite a huge amount of data that is not of much relevance, the Twitter stream contains helpful data [16]. Numerous current news items have been reported utilizing Twitter directly from clients at the site [17]; Examples being US Airways flight 1549 arriving in the Hudson River [18] or road riots amid Iran's 2009 presidential decisions [19]. Since tweets are regularly sent from handheld devices, they pass on more promptness than other interpersonal interaction methods.

In the current years, there has been a considerable amount of research in the area of online networking information. There is a restricted report towards the general wellbeing data framework. A few authors [20] worked in the field of general public health issues utilizing web-based social networking information.

Authors in [9, 10] proposed a technique amid the H1N1 pandemic utilizing Twitter-based data with particular keywords. Other authors likewise utilized some different procedures or techniques for discovering queries and keywords like Google web search queries connected with flu epidemic. The clustering of Twitter messages by subject and extraction of important intelligible marks for each group is utilized by [21]. Some researchers utilized content examination and relapse models to quantify and screen open concern and levels of infection amid the H1N1 pandemic in the United States [22]. Some different sicknesses like [19] cholera is additionally explored to discover the cholera flare-up. The creators additionally built up a system that gave a measure of clients influenced by flu (swine influenza) within a group or gathering. Machine learning procedures like SVM have been used in [23]. Flu pandemic has been tracked by monitoring the social web in [20]. Tweets related to dental pain have been analyzed in which may help dentists to disseminate dental health-related information. 1000 tweets related to use of antibiotics have been analyzed in [7] to discover sign of misinterpretation or misapplication of antibiotics. 60,000 tweets about cardiac arrest and resuscitation collected during 38 days via a set of 7 search terms have been analyzed in [24]. Twitter data is very useful in the study of a wide range of health-related problems [25–28]. The unstructured data obtained

from Twitter has to be preprocessed before it can be used for any type of analysis. Authors in [29] have done text normalization by using Deep Neural Network approach. Predictive modeling has been used by authors in [30] for prediction of cases, deaths, and recoveries due to COVID 19.

## 3 Analysis Techniques for Twitter Data

Data extracted from Twitter can be used for analysis and prediction in various disciplines such as medicine, engineering, biology economics, forensics, etc. [31]. Various aspects can be measured from the tweets based upon the aims and objectives of the work being carried out. Tweets could be analyzed for carrying out sentiment analysis, for the detection of objects, for prediction or may be for content analysis [1]. Content analysis and sentiment analysis deals with text analysis of the tweets to arrive at some useful results. In content analysis, the text which is extracted is generally a word related to the area of study which is used for some sort of prediction [13, 32], whereas in sentiment analysis the words extracted generally represent the emotions of the people [8]. Several tools can be used for sentiment analysis. They can be broadly classified as lexicon-based tools and machine learning based tools. A "dictionary" of words with their sentimental meanings is required in lexicon-based tools so that the keywords can be categorized as positive, negative or neutral [33]. Two frequently used dictionaries are Linguistic Inquiry and Word Count (LIWC) [34] and Affective Norms for English Words list (ANEW) [35]. This lexicon-based approach is simple to implement, but suffers from low recall if the size of lexicon is very small or if very few numbers of lexicon words are present in the text. Another limitation is that this method can classify the sentiment only into three categories: positive, negative, and neutral; but in the real-world sentiments other than these extremes also exist.

Machine learning approaches for text analysis are probabilistic in nature. In this approach initially, a training dataset is provided which consists of sample dataset with known positive or negative sentiment. This dataset is first used to develop a model and then this model is further applied for text analysis of the new data items that are the tweets. WEKA, Microsoft NLP Toolkit, and SciKitLearn are some of the popular ML-based sentiment analysis tools.

## 4 Proposed Approach

The proposed strategy analyzes the data from web search queries and tweets. The first step here is the recognition of queries and tweets that are pertinent to the presence of influenza, influenza symptoms, and medication for influenza. Some important and relevant keywords from tweets used in this work are: Influenza, Swine flu, H1N1, Decongestants, Rapivab, Tamiflu, Relenza, Flomist, Zanamivir, etc. These tweets have been used to find out that how many people use Twitter in different countries

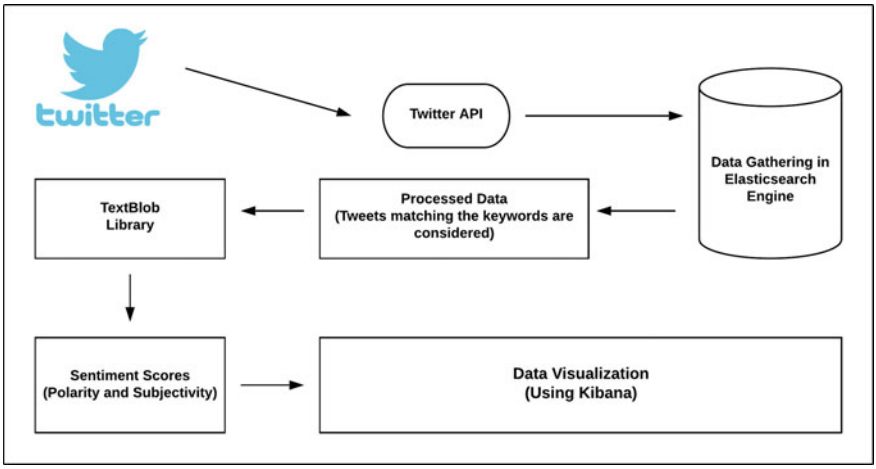



**Fig. 1** Proposed approach

to share their experiences on the same. This helps to find out how these diseases are being managed by people without going to the hospital and how it can be useful for other users. Later, using IDSP data and Twitter data a comparison has been carried out for the analysis of the number of cases in both the scenario. This might be of great help to analyze how social networking sites help the patients to encounter the starting symptoms of the disease and help to cure it. Based on this study the average number of cases of the disease that have been resolved in India without going to the hospital has been estimated. Also, the analysis of where India stands when it comes to Twitter data and what is the scenario of Influenza in other countries has also been studied and the visualization of the results obtained is presented.

Sentiment analysis is a task of NLP which is the subfield of AI (Artificial Intelligence) which is used to help the machines to deal with the human languages. Naive Bayes and Decision Tree are the best machine learning algorithms and are used in this work too. Firstly, perfect data is collected from Twitter for training and testing, this is done using the keywords which are mentioned in the python code which include Influenza, H1N1, Decongestants, Rapivab, Tamiflu, Relenza, etc. Secondly, Textblob is used, it is a python package that is used to do a lot of NLP tasks. It is used for polarity and subjectivity. Polarity lies between the range of $[-1, 1]$, discussed later in Sect. 5 and subjectivity refers to the personal opinion, emotion, or judgment. In the third step, the extracted tweets were stored in Elasticsearch Engine in the JSON (JavaScript Object Notation) format for better results as the data is a lot, which needs to be maintained in the proper format for better visualization. Also, as data is very huge and in different languages, only English language tweets were taken out of all the tweets. Later in the fourth step, the stored data was then visualized in Kibana, which is an open-source data visualization and exploration tool. Kibana is used for searching, viewing, and visualizing data indexed in the Elasticsearch and later for

**Table 1** Hashtags

| | | |
|---|---|---|
| #cold | #H1N1 | #Decongestants |
| #Flu | #Influenza | #influenzavaccine |
| #virus | #flushot | #influenzab |

analyzing the data via pie charts, histograms, bar charts, tables, and maps. Figure 1. Explains the steps of the proposed approach in a flow diagram.

A detailed explanation of the various tools and techniques is given below.

### *4.1 Data Gathering*

Tweepy API has been used here to collect the tweets related to H1N1 in order to understand the opinion of Twitter users about this disease. In order to do this firstly an account was created on Tweepy API and this was then linked to our Twitter account. 33,120 tweets were extracted in this manner and they were stored in a database. Some of the hashtags used to retrieve the tweets are given in Table 1.

### *4.2 Data Preprocessing*

The first step in preprocessing was removal of all the irrelevant data such as stop words, URL's, special characters, symbols white spaces, etc. The second step was removal of all the words which were not English language words, as the current research work is working only for English words.

### *4.3 Generation of Polarity*

The data obtained from the Tweets are an unstructured data. The words present in these Tweets is used to determine its polarity which has numerical value associated with it between −1 and 1. Negative value implies negative sentiment, positive value implies positive sentiment, and a value of zero means neutral sentiment. Subjectivity is another parameter which has its value in the range of 0–1. A value near 0 means that the tweet is based on factual data, whereas a value near to 1 means that it is mostly based upon public opinion. Textblob Python library has been used in this work to calculate the polarity and subjectivity.

### *4.4 Machine Learning Model*

The final step involved is classification using Machine Learning approach. Naïve Bayes and decision tree supervised learning techniques have been used here. The pre-processed dataset was used to build a model using tenfold cross-validation for the two classifiers and the performance of the two was measured using accuracy as the parameter.

After prediction using the above approach, the results were compared with the data available at IDSP. The IDSP data is easily accessible from the 'Ministry of Health and Family Welfare-Government of India' site, i.e., www.idsp.nic.in. The key motive of this web portal is to maintain and strengthen the decentralized laboratory-based IT facilitated disease surveillance for the rampant prone disease to observe the trends of the disease and to detect and respond to the epidemics in the initial stage via Rapid Response Team (RRTs). This data is updated every month and provides previous year data also in the pdf format. In this work, first this data has been converted into CSV format using python code and then has been visualized using POWER BI. In both the cases it has been observed that there is a huge difference between the number of cases encountered in twitter data and IDSP data.

## 5 Results and Discussions

This work is done to predict Influenza, when a person is suffering from Influenza or who has recovered, let's say posts something on Twitter which might include the symptoms, precautions, or the medication then the tweet is very useful. If something bad is written in the tweet it shows the tweet is negative polarity and a person is/was infected. If something good is written then it shows the tweet is positive and a person has recovered from the illness and the tweet might include the precautions, cure, and medication also.

The program was run for approximately 30–32 h which means one and a half-day and the number of tweets/cases encountered were 33,120 in total. The number of Tweets remaining after preprocessing was 14,530. The polarity was sub-divided into four ranges and the number of tweets encountered in these four ranges is shown in Table 2.

**Table 2** Sentiment assignment to various tweets

| Polarity Range | Count |
|---|---|
| −1 to −0.5 | 1273 (8.95%) |
| −0.51 to 0 | 5461 (38.38%) |
| 0.1–0.5 | 6331 (44.49%) |
| 0.51–1.0 | 1165 (8.19%) |

**Table 3** Number of cases (approx.) encountered per day using Twitter and IDSP data

| S. No. | Cases per day by Twitter data | Cases per day by IDSP data |
|---|---|---|
| 1 | 43 | 21 |

The performance of the classifiers was evaluated and it was found that Naïve Bayes achieved an accuracy of 89.36% and Decision Tree achieved an accuracy of 92.49%. Hence, it can be said that Decision Tree has performed better than Naïve Bayes.

Furthermore, out of the total tweets 65 of them were encountered in India. Based on this result the average number of cases per day has been estimated as 43 using the sentiment analysis approach. As per IDSP data, the number of cases from 1st January 2018 to 22nd April 2018 (81 days) was 1678. This results in an average of 21 cases per day.

It can be seen from here that there is huge difference between the results obtained from both method. This is depicted in Table 3.

## *5.1 Visualization of Twitter Data Using Kibana*

Figure 2 shows the polarity ranges of the tweets within the range of –1 to 1. Which is further sub-divided into 4 ranges like between –1 and –0.5, –0.51 to 0, 0.1–0.5 and 0.51–1. But here, the proposed research work has intended to use 2 ranges –1 to 0 and 0.1–1. As the analysis is about negative and positive tweets only.

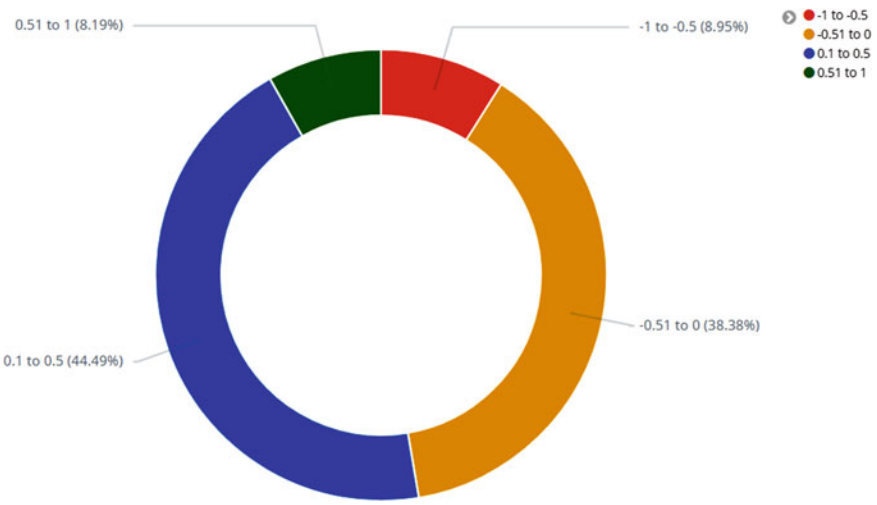



**Fig. 2** Number of Counts in the polarity range table

The ranges are the sentiments obtained from the tweets. For more accuracy, the data has been classified into four ranges here. The sentiments here are classified based on the views of the people in the tweets.

Figure 3 shows the number of counts based on the locations where the tweets were posted. The counts are in the tabular form and the data is further visualized on the graph.

Figure 4 shows location-wise Bar Graph depicting the number of counts based on the location world-wide. As it can be seen here the number of counts in India is 65 and is on 21st position.

Table Request Response Statistics

| Location | Count |
|---|---|
| United States | 317 |
| FART | 231 |
| London | 182 |
| London, England | 157 |
| Los Angeles, CA | 142 |
| New York, NY | 124 |
| England, United Kingdom | 105 |
| Lagos, Nigeria | 102 |
| Chicago, IL | 101 |
| United Kingdom | 101 |
| UK | 100 |
| Washington, DC | 100 |
| California, USA | 97 |
| Canada | 97 |
| Nigeria | 91 |
| Houston, TX | 80 |
| USA | 79 |
| France | 72 |
| Florida, USA | 69 |
| Seattle, WA | 66 |
| India | 65 |
| Malaysia | 65 |
| Atlanta, GA | 61 |

**Fig. 3** Number of counts based on the location

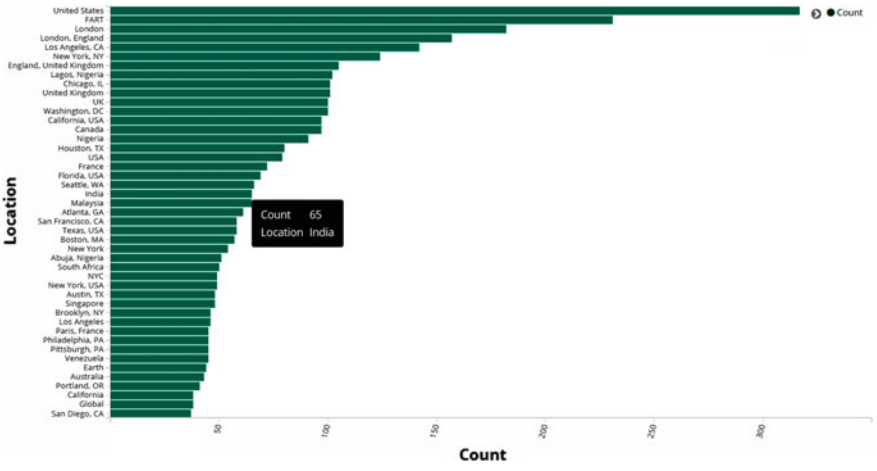



**Fig. 4** Location wise bar graph

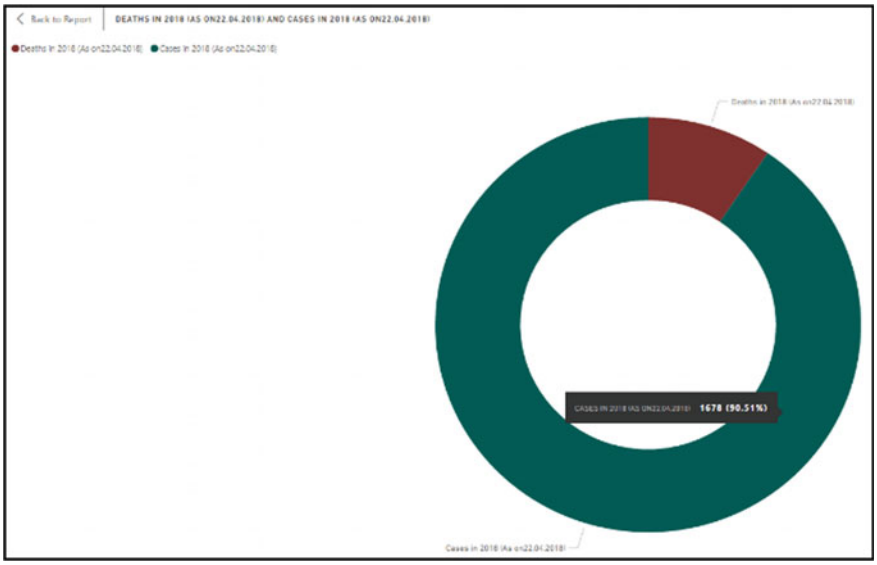



**Fig. 5** Overview of number of total cases to number of total deaths in INDIA

## *5.2 Visualization of IDSP Data Using Kibana*

Figure 5 depicts the number of total cases to the number of total deaths in the period of 81 days in India. The cases registered under IDSP were found to be 1678 and number of total deaths was 176. The data was of 29 states and 7 union territories from 1st January 2018 to 22nd April 2018.

Figure 6 shows the number of cases state/UT wise. In Telangana state it has 16 cases which is 0.95% of total cases as on 22nd April 2018. The maximum number of cases were encountered in Rajasthan. This shows that the one of the most affected parts of India when it comes to influenza is Rajasthan. The number of cases encountered in Rajasthan is 1343 but analysis of Twitter data shows that there would be a greater number of cases in the future. To overcome this a strict step should be taken to minimize the problem in the coming days.

Figure 7 shows the number of cases state/UT wise. In Uttar Pradesh state it is seen it has 3 death cases which are 1.7% of total deaths as on 22nd April 2018. Again, the number of death cases in Rajasthan is maximum as the cases encountered here were maximum. It's a matter of serious concern for India. The facilities should be provided, and medications should be provided as soon as possible to reduce this number.

Figure 8 Depicts the results presented in Table 1 in the form of a graph. The results obtained from the analysis show that there is a huge difference between IDSP data and Twitter data. Results obtained from Twitter data show approx. double the number of cases as reported by IDSP. The difference depicts how the Twitter data

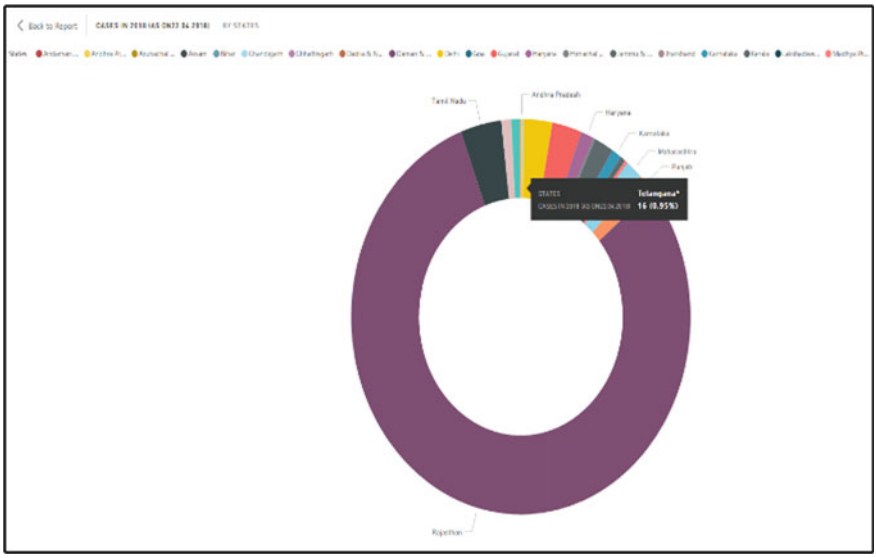



**Fig. 6** State/UT wise count of cases

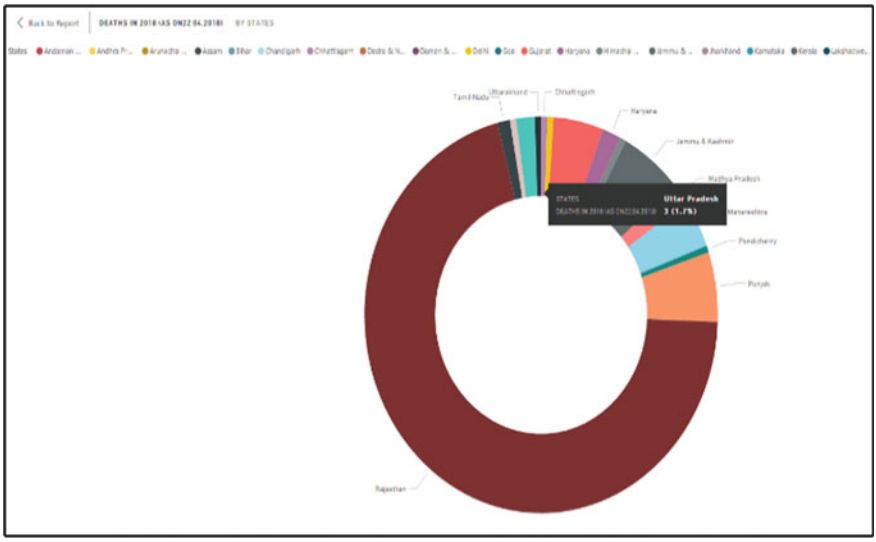



**Fig. 7** State-wise count of deaths

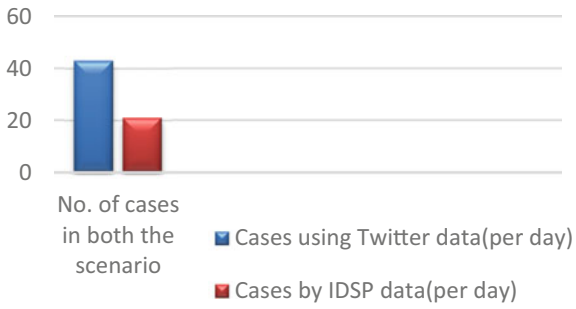



**Fig. 8** Number of cases encountered in Twitter data and IDSP data

can be beneficial in providing quick review and feedback on influenza-like illness without going to hospitals. Similarly, if anyone had posted anything about COVID-19 in early stages there is a possibility that it could have been very beneficial in taking preventive measures more efficiently and in making medication for it

## 6 Conclusion

Twitter offers a platform for various discussions and opens doors for observation of general wellbeing of the people. In this work, people's opinions were studied about H1N1. The sentiment analysis of Twitter data was carried out and the accuracy obtained using Naïve Bayes and Decision Tree was 89.36% and 92.49%, respectively. A comparison of the IDSP data for seasonal influenza in India from 1st January 2018 to 22nd April 2018 with the twitter data is done. The results obtained here can give quick feedback or reviews on side effects for swine influenza, prevention strategies used by the users and their reviews about the drugs or medications they took. This data and result could be used for taking preventive measures from the disease prediction by government and health organizations by giving more supply in a particular region which will enhance their business too. Also, the outcome supports that the utilization of online networking for the following disease will be used for knowing the public health in the society more accurately and will be able to resolve the problem in a shorter period.

## 7 Limitations and Future Scope

The results obtained in this work are very interesting and encouraging. As mentioned, web was used and found the views of the people and their perceptions for certain diseases via their tweets. Tweets were extracted which matched to the defined keywords. But in different regional languages the flu may be called something else and the users might have mentioned the disease in their own language. As this research

did not look for those keywords therefore those tweets were not extracted, so this research is limited for language English only. Another limitation is, the tweets are limited to specific length and that is maximum 140 words. The tweets don't have relevant information in 140 words, the data in the tweet contains very less facts/figures and doesn't offer much data to work upon.

As mentioned above two limitations were observed. One limitation was regarding the extraction of the data from the Twitter which is being posted in some other language which doesn't match the mentioned keywords. So, one of the future scopes of the work is to have more accurate and relevant data, and for that one needs to have a surveillance system that detects such tweets also, and one can look forward in the future to work upon this for better results. The second limitation was about how the tweets are limited to 140 characters and are not relevant enough for the more accurate analysis purpose. So, another future scope is, the detections could be improved with simple grammatical analysis along with other recognized classification techniques. Last year, Twitter also said they will be extending the length of the characters in the tweet. That might be very helpful to work upon and extract more relevant data.

## References


1. Poecze F, Ebster C, Strauss C (2018) Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. Proc Comput Sci 130:660–666
2. McCalman J, Bainbridge R, Brown C, Tsey K, Clarke A (2018) The aboriginal australian family wellbeing program: a historical analysis of the conditions that enabled its spread. Front Public Heal 6:26
3. Amato PR (2010) Research on divorce: continuing trends and new developments. J Marriage Fam 72(3):650–666. https://doi.org/10.1111/j.1741-3737.2010.00723.x
4. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: Real-time event detection by social sensors. In: Proceedings of the 19th international conference on world wide web, WWW'10, pp 851–860. https://doi.org/10.1145/1772690.1772777
5. Prier KW, Smith MS, Giraud-Carrier C, Hanson CL (2011) Identifying health-related topics on Twitter. In: International conference on social computing, behavioral-cultural modeling, and prediction, pp 18–25. https://doi.org/10.1007/978-3-642-19656-0_4
6. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH (2013) Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 15(8):e177. https://doi.org/10.2196/jmir.2775
7. Bechmann A, Lomborg S (2013) Dissemination of health information through social networks: Twitter and antibiotics. New Media Soc 15(5):765–781. https://doi.org/10.1016/j.ajic.2009.11.004
8. Malik M, Habib S, Agarwal P (2018) A novel approach to web-based review analysis using opinion mining. Proc Comput Sci 132:1202–1209
9. Freberg K, Palenchar MJ, Veil SR (2013) Managing and sharing H1N1 crisis information using social media bookmarking services. Public Relat Rev 39(3):178–184. https://doi.org/10.1016/j.pubrev.2013.02.007
10. Jania VK, Kuma S (2015) An effective approach to track levels of influenza-A (H1N1) pandemic in India. Proc Comput Sci 70:801–807
11. Yaqub U, Chun SA, Atluri V, Vaidya J (2017) Analysis of political discourse on twitter in the context of the 2016 US presidential elections. Gov Inf Q 34(4):613–626. https://doi.org/10.1016/j.giq.2017.11.001

12. Leitch D, Sherif M (2017) Twitter mood, CEO succession announcements and stock returns. J Comput Sci 21:1–10
13. Wang W, Chen L, Thirunarayan K, Sheth AP (2012) Harnessing twitter 'big data' for automatic emotion identification. In: Proceedings - 2012 ASE/IEEE international conference on privacy, security, risk and trust and 2012 ASE/IEEE international conference on social computing, SocialCom/PASSAT 2012, pp 587–592. https://doi.org/10.1109/SocialCom-PASSAT.2012.119
14. Malik M, Naaz S, Ansari IR (2018) Sentiment analysis of Twitter data using big data tools and Hadoop ecosystem. In: International conference on ISMAC in computational vision and bio-engineering, pp 857–863
15. Chaudhary S, Naaz S (2017) Use of big data in computational epidemiology for public health surveillance. In: 2017 international conference on computing and communication technologies for smart nation, IC3TSN 2017, 2018, Oct 2017. https://doi.org/10.1109/IC3TSN.2017.8284467
16. Bifet A, Frank E (2010) Sentiment knowledge discovery in Twitter streaming data. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 6332 LNAI, pp 1–15. https://doi.org/10.1007/978-3-642-16184-1_1
17. Phelan O, McCarthy K, Smyth B (2009) Using twitter to recommend real-time topical news. In: RecSys'09—proceedings of the 3rd ACM conference on recommender systems, pp 385–388. https://doi.org/10.1145/1639714.1639794
18. Heppermann C (2013) Twitter: the company and its founders. ABDO
19. Chunara R, Andrews JR, Brownstein JS (2012) Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 86(1):39–45
20. Lampos CNV (2010) Tracking the flu pandemic by monitoring the social web. In: 2010 2nd international workshop on cognitive information processing (CIP). IEEE Computer Society, pp 411–416
21. Basha PS Document based clustering for detecting events in microblogging websites
22. Siston AM et al (2010) Pandemic 2009 influenza A(H1N1) virus illness among pregnant women in the United States. JAMA J Am Med Assoc 303(15):1517–1525. https://doi.org/10.1001/jama.2010.479
23. Aramaki E, Maskawa S, Morita M (2011) Twitter catches the flu: detecting influenza epidemics using Twitter. In: Proceedings of the conference on empirical methods in natural language processing, pp 1568–1576
24. Bosley JC et al (2013) Decoding twitter: Surveillance and trends for cardiac arrest and resuscitation communication. Resuscitation 84(2):206–212. https://doi.org/10.1016/j.resuscitation.2012.10.017
25. Zhang L, Hall M, Bastola D (2018) Utilizing Twitter data for analysis of chemotherapy. Int J Med Inform 120:92–100. https://doi.org/10.1016/j.ijmedinf.2018.10.002
26. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ (2017) Forecasting the onset and course of mental illness with Twitter data. Sci Rep 7(1):1–11. https://doi.org/10.1038/s41598-017-12961-9
27. Jain VK, Kumar S (2018) Effective surveillance and predictive mapping of mosquito-borne diseases using social media. J Comput Sci 25:406–415. https://doi.org/10.1016/j.jocs.2017.07.003
28. Gohil S, Vuik S, Darzi A (2018) Sentiment analysis of health care tweets: review of the methods used. J Med Internet Res 20(4):e43. https://doi.org/10.2196/publichealth.5789
29. Arora M, Kansal V (2019) Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis. Soc Netw Anal Min 9(1):12
30. Hamzah FAB et al (2020) CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. Bull World Heal Org 1:32

31. Wang X, Gerber MS, Brown DE (2012) Automatic crime prediction using events extracted from twitter posts. In: International conference on social computing, behavioral-cultural modeling, and prediction, pp 231–238
32. Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS ONE 6(5):e19467
33. Zhang L, Ghosh R, Dekhil M, Hsu M, Liu B (2011) Combining lexicon-based and learning-based methods for Twitter sentiment analysis. HP Lab Tech Rep HPL-2011 89
34. Pennebaker JW, Booth RJ, Francis ME (2007) Linguistic inquiry and word count: LIWC [computer software], vol 135. Austin, TX liwc.net
35. Nielsen FÅ (2011) A new ANEW: evaluation of a word list for sentiment analysis in microblogs. arXiv:1103.2903

# Review of Denoising Framework for Efficient Removal of Noise from 3D Images


**Anand B. Deshmukh and Sanjay V. Dudul**

**Abstract** An image is a distributed amplitude of colors on a plane. An image may be in the form of two-dimensional image or three-dimensional image. Such images are compiled using optical sensors like camera and are processed using various image processing tools for better visualization. Purpose of the image processing is not limited for better visualization, but it is extended to remove noise from the captured image. Noise is a random variation of brightness, contrast and color pallets in an image. In the present discussion through review of denoising framework for efficient removal of noise from 3D images, different filters which are used so far for removal of noise are discussed. The research work is further extended by designing novel denoising framework for efficient removal of noise from the 3D image.

**Keywords** Denoising · Gaussian filter · Impulse filter · Spectral filter · Rician filter · Synthetic image · Gray scale image and color image

## 1 Introduction

An image is collection of number of rows and columns of pixels. Pixel is the smallest element in the image. They are represented by 1's or 0's in binary image or small integer value if it is gray image or color image. Accordingly images are two-dimensional and three-dimensional.

Images are either acquired using optical sensors like camera during which noise signal gets added in an image. On the other side, the images are transferred through possible media like wired or wireless transmission using different communication protocol which again causes noise introduction. Introduction of noise signal causes degradation in quality of an image. To recover maximum original information in an image, it is significant to use effective methodology for removing noise signal from an image. Technique used to remove unwanted signal from an image is known as denoising technique.

A. B. Deshmukh (✉) · S. V. Dudul
Sant Gadge Baba Amravati University, Amravati, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_32

There are different types of noise added into an image with different level. In the present research work, exhaustive review is carried out on various noises such as speckle noise, impulse noise, Gaussian noise and Rician noise along with an innovative approach for noise removal from an image is disclosed.

## *1.1 Understanding the Noise*

Noise is generally represented as a random variation of color and brightness information in an image [1]. There are different sources of noise from which it can be added into an image like in case while acquiring through optical sensors like camera or scanner. Other possible source of noise in an image is at the time of processing an image for transmission like compressing, encoding and other significant steps or at the time of transmission of an image using a different communication protocol [2].

Noise signal exists in different levels. Noise can range from almost imperceptible pixel spot to a radio astronomical image which remains completely noisy from which very small amount of information or even sometimes no information can be retrieved.

**Gaussian Noise**

The Gaussian noise arises mostly during image acquisition. Significant reasons for Gaussian noise are heat, random illumination and electronic circuit noise. Primarily the optical sensors used for image acquisition tends to heat up due to the surrounding temperature and due to the amount of current they draw from battery source for own operation, this causes loss of information in an image. Poor illuminating conditions at the time of image capturing cause inherent noise in an image. The third significant reason is that electronic circuit noise. It is caused due to internal and external sources. External sources include radio frequency transmitters and receivers, nearby conductors, ignition system and many other possibilities. Internal reasons include faulty components, age of the component like lens and many other significant things [3].

Gaussian noise is uniformly distributed over an image. It causes each pixel on an image to be a sum of pixel value and Gaussian noise signal value. Mathematically, Gaussian noise can be represented as

$$\mathcal{P}G(z) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

where

$\mathcal{P}$ is the probability density function,
$z$ represents the gray level,
$\mu$ mean value,
$\sigma$ standard deviation [4].

**Impulse Noise**

An image is captured using optical sensors. When light signal misaligns with the sensors it creates a noise. Whenever an image is transmitted on communication channel, due to the external noise, image becomes noisy. Such type of noise is called as impulse noise. Due to imperfection in converter designs might introduce impulse noises [5]. Impulse noise can be seen unsubstantially as white spots of pixel in dark area of the image or dark spot of pixel in bright area of the image. Whenever a positive spike arrives having value greater than the value of surrounding pixels then that pixel appears as white spot. On the other hand, if the noise spike has value smaller than the value of surrounding pixel, then that pixel appears as dark spot [6]. For effective removal of impulse noise median filter [7] or morphological filter, [8] can be used.

Mathematically, impulse noise can be represented by

$$y_{ij} = \begin{cases} \text{smallest impulse vale or highest impluse value} \\ (x_{ij}) \end{cases}$$

where

$y_{ij}$ noisy image pixel,
$x_{ij}$ uncorrupted image pixel [9].

**Speckle Noise**

Noise is random variation in pixel values, and this adjust the actual content of the image. Addition of such pixels in an image leads to form a noisy image. Out of the different categories of the noises, the noise which occurs due to environmental conditions is speckle noise. It occurs at the time of procurement while capturing the image [10]. Speckle noise is generally observed in medical images and RADAR images.

Mathematically, speckle noise can be represented as,

$$G = e + n * e$$

where

$G$ speckle noise,
$e$ input image,
$n$ uniform image noise with mean and variance [11].

**Rician Noise**

Rician noise occurs in low signal-to-noise ration regions and it is a signal dependent noise. Rician distribution occurs when an image is exposed to resonance. Due to this, random fluctuations are observed. It also causes signal dependent which may reduce the contrast of an image. Wavelet and other different methodologies and algorithms are used to eliminate the Rician noise.

## 2 Recapitulating Image Denoising

Noise is the random variation in amplitude of pixel in an image, as compared with the neighborhood pixel values. This corrupts the information which an image is carrying with it. For optimized working and perfect investigation of an image, true image quality with zero noise inclusion is required. The process of extracting the false contents in an image to restore the original image quality is called as denoising. Various authors have reported various novel algorithms, filtering techniques and methodologies for denoising.

**Gaussian Denoising**
Manyu Wang and Shen Zeng proposed an image denoising method based on Gaussian filter. The proposed algorithm learns from the weighted average thoughts of particle filter. In the novel technique of noise removal instead of considering only gray information of the image, gray information along with image structure information is also considered for denoising which gives enhanced denoising [12]. Figure 1 discloses the outcome of the proposed method, as reported by the authors.

Tanzila Rahman et al. proposed a reduced the Gaussian noise using fuzzy filter in their research work. A $3 \times 3$ filtering window with 8 neighboring pixels are used to calculate the input image. In this to calculate the percentage of corruption of an image, the input image is divided by 8. Because of this, 8 neighboring pixels can be easily calculated. For color image instead of directly processing, the individual color

**Fig. 1** Experimental results for the proposed method

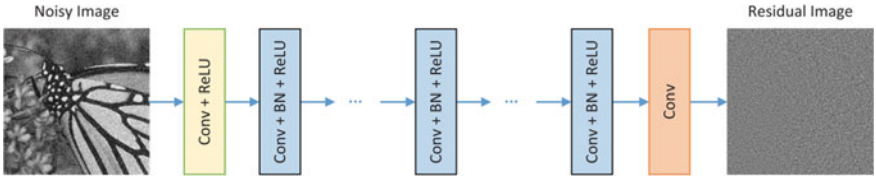



**Fig. 2** Architecture of the proposed system

components R, G and B are processed separately and finally concatenated to get the final output image [13].

Kun He et al. proposed a technique to remove the Gaussian noise completely than to reducing it. Further, in the proposed technique, the pixels are categorized based on the local continuous smoothing in the image, into three classes, on noise point, on noise edge and in the local texture. At the second stage, according to the local continuity and the texture property, edge information and the texture information are extracted. This helps to locate the noise point precisely. In the final stage of denoising Gaussian noise, adaptive neighborhood is used to eliminate the noise point [14].

Kai Zhang et al. investigated the development of feed forward denoising CNN model to remove the noise. Improvised training process and denoising performance residual learning and batch normalization techniques are adopted. Through the proposed algorithm removes the buried clean image due to which it can denoising super-resolution and deblocking can be handled. Further authors claimed to have architecture capable of blind Gaussian denoising [15] (Fig. 2).

Florian Luisier, Thierry Blu and Michael Unser proposed an optimal methodology for thresholding algorithms to remove the Gaussian and Poisson noise. In this methodology, the denoising process is considered as linear expansion of the threshold which is optimized by mean squared which is obtained through non-Bayesian approach [16].

Muthukumar et al. propose a restoration algorithm for color image suffering from Gaussian noise and impulse noise. In this technique, median filters with architectural data are used to identify impulse noise pixel. Gaussian noise is removed by utilizing the total variation minimization algorithm. The proposed method is claimed as effective technique for restoring an image while retaining the structural information and visual details [17].

Takayuki Hara and Haike Guan disclose an algorithm for denoising color image without degrading the information content in an image at promising speed. Research work utilized maximum posterior prediction is used which is based on Gaussian model. The correlation in the color components is reduced along with computational complexity due to non-iterative filtering and matrix operations [18].

Dinh-Hoan Trinh et al. reported the issues due to Gaussian and Poisson noise in medical images and introduced patch-based learning approach for noise removal. Regression model is used for denoising which removes noise based on nearest neighbors of an infected pixel group. A set of k-nearest neighbors is used to learn the

regression model and performs denoising effectively in highly noise affected images [19].

Mrs. C. Maithili and Dr. V. Kavitha suggested a filtering technique by considering three primary colors to be filtered separately followed by gain adjuster unit to compensate the overall gain and then combine the three module outputs together to form an image [20].

**Impulse Denoising**

Hemant Kumar Aggarwal and Angshul Majumdar disclosed denoising technique for hyperspectral images. The optimization problem is framed for the noise function in the hyperspectral images and it is removed using spatial correlation between the images. The other optimization issues are resolved using Split Bregman technique, and the denoised hyperspectral image is claimed to be superior than previously reported techniques [21].

Hakran Kim, Seongjai Kim proposed diffusion-based denoising method to deal with impulse noise. A novel impulse moving anisotropic diffusion filter is designed which regulates the impulses and max and mins without disturbing neighboring pixel amplitudes [22].

Nasar Iqbal, Kashif Ahmad and Waleed Shah Jehan disclosed a technique to denoise an image using adaptive median filtering in which the infected pixel is replaced by median value of $3 \times 3$ window. Technically, the clean pixels are used to calculate the median values and it is used to replace the noisy neighbor pixel. Outcomes are claimed to give optimized mean square error, peak signal-to-noise ratio and universal quality index [23]. The subsequent Fig. 3 discloses the relation between PSNR and noise density which is reported in the research work experimental analysis.

Bo Xiong, Zhouping Yin discussed universal denoising framework based on nonlocal means filtering. Denoising is pursued in detection and filtering stages. For detection of the impulse infected pixel, outlyingness ratio is used which is followed by dividing all pixels into four clusters depending on outlyingness ration value. Subsequently, to precisely detect the infected pixels iterative model and fine coarse to fine strategy are utilized. To filter the impulses, a reference image is introduced to extend the NL means. Outlyingness ratio and NL means are combined and proposed as a denoising and restoration model. The proposed technique claimed to achieves high peak SNR and improved image quality by efficiently removing impulse, Gaussian and mixed noise [24].

Snigdha Tariyal et al. utilized interband spectral correlation along with intraband spatial redundancy in their research work in transform domain as a thin representation. Set of transform coefficients are used to represent the intraband spatial redundancy and rank deficient matrix is used to model the interband spectral correlation. Proposed work achieved better optimization performance by utilizing Split Bregman technique [25].

Dev R. Newlin, C. Seldev Christopher discussed denoising technique for impulse noise affected image through adaptive weighted interpolation along with edge sensing

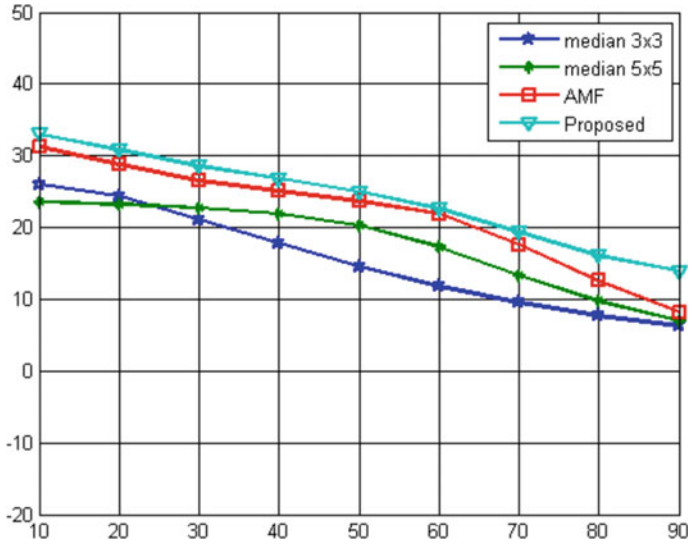



**Fig. 3** Relation of PSNR and noise density

and adaptive window size. The claimed architecture promises better restoration when the performances are compared with existing filtering models [26].

S. Indu and Chaveli Ramesh denoise an image using image fusion technique. In this technique, multiple images are fused together while retaining the information contents of the individual image. To denoise an impulse, the uncorrupted pixels of a noisy image from different sensors are combined together [27].

Angshul Majumdar et al. address the issue in noise affected hyperspectral images. Joint-sparsity promoting dictionary and blind compressed sensing framework dictionary is suggested in the research work for denoising [28].

**Speckle Denoising**

Meenal Gupta and Amit Garg discovered a technique to preserve the edges and to minimize the speckle noise. The technique reported in the research work is built using detail preserving anisotropic diffusion and optimized Bayesian non-local mean filters. The outputs of the filters are combined using homogeneity map, which differentiates homogeneous and edge region of the image. Through the proposed technique improved SNR and edge keeping index are claimed [29].

Latha Parthiban and R. Subramanian uncover speckle denoising using contourlets to attain improved SNR performance. Contourlet transform is the two-dimensional transform developed for image processing in discrete and continuous domain. This algorithm helps to restore the images to the level that makes diagnostically relevant image content. Using the proposed technique, superior performance in terms of speckle suppression and edge preservation is achieved [30].

Amit Garg and Vineet Khandelwal discovered a method for speckle denoising using combined approach of bilateral filter and detail preserving anisotropic diffusion. The local coefficient of dispersion which is estimated through flexible window is used to frame a binary classifier map in the research work. The superiority of the proposed technique is modeled around PSNR, SNR and EKI parameters [31].

Celia Barcelos, Luciana Vieira disclose a method which identifies the noise pixel and its location using a function which controls the smoothness velocity and adjusts the fidelity [32].

Mohammad Rahman, Mithun Kumar, B. Borucki, K. S. Nowinski uncover a novel technique for speckle denoising of ultrasound images. The proposed technique uses vector triangular formula as an extra energy reduction function. In this technique, the energy of pixel is considered for identifying noisy pixels, if the energy of pixel which is calculated using vector triangular formula. Noise pixels are observed if the calculated pixel is less than that pixel in terms of distance energy. By optimizing the additional energy for functional operation the noise pixel energy is balanced [33].

Tsuyoshi Murakami, Koichi Ogawa disclose a technique for speckle denoising of optical coherence tomography (OCT) images. In this technique, wavelet transform is used for image restoration. The signal-to-noise ratio is claimed to be increased to 25 dB [34].

Somkait Udomhunsakul and Pichet Wongsita disclose a method for restoring an image infected from speckle noise using combine approach of Weiner filter theory and wavelet transform while maintaining the information contents. In this technique, 2D discrete transform of logarithmic image is calculated followed by Weiner filtering over sub-band areas. Subsequently, inverse wavelet transform and inverse logarithm are computed. For parameter evaluation, MSE and EP are used [35].

Raul Malu Gan, Romulus Terebeu, Christian Germain, Monica Bordal, Mihaela Ciulariu proposed classical signal processing technique that is independent component analysis method for speckle denoising. In this technique, sparse code shrinkage algorithm based on independent component analysis is applied. Inverse of the unknown mixing matrix is estimated using fast ICA algorithm. This is followed by application of shrinkage operator for each determined component [36].

Celia Barcelos, Luciana Vieira reveal the use of variational method to suppress speckle noise. In this technique, location of noisy pixel is detected by a function controlling the smoothness and adjusts the fidelity. The proposed model claimed to be faster and perfectly remove the speckle noise while retaining the image quality [37].

**Rician Denoising**

Debajyoti Misra, Subhojit Sarker, Supriya Dhabal and Ankur Ganguly disclose the use of genetic algorithms for the use of denoising an image suffering Rician noise. The speckle noise is arising mainly due to low SNR, and genetic algorithm is used in the proposed rectifies the real and imaginary data before magnitude image construction. Crossover and adaptive mutation are combined in the experimental model which attains better convergence and robustness [38].

Robert D. Nowakuncover's wavelet domain filtering method to denoise Rician noise from in image addressing to the random variations in the signal and noise. Through this approach, random fluctuations in the contrast occurrences in an image can be regulated efficiently [39].

Herng-Hua Chang et al. disclose a post-acquisition denoising algorithm to regulate the random fluctuations and suppress bias introduced due to the Rician noise. The proposed filtering technique considers median, radiometric and geometric components to replace the intensity values of neighboring pixels. An entropy function is associated with the neighboring pixels along with parameter automation scheme to balance the interferences through FMFs [40].

Efstathios Sotiriou et al. report the issues due to magnetic resonance images which swiftly outperforms Rician denoising high resolution images. This approach is attained by designing hardware architecture of the proposed algorithm. The images are segmented and processed in pipeline. Due to this novel approach, the non-infected pixels can be bypassed from the processing steps [44].

Xinji Liu and Tao Dai proposed principal component analysis (PCA) with local pixel grouping (LPG)-based denoising model focusing on low noise levels of image. Authors have introduced a guide image and addressed the issues of robustness and high noise through PCA transform estimation. Block matching is performed between noise image and guide image for PCA transform estimation [41].

Erik Bochinski et al. proposed a sparse image representation model through regularized gradient decent training of steered mixture of experts. Instead of Gaussian mixture model trained by expectation maximization, gradient descent optimization is used as primary training object in the proposed model provided better performance in terms of sparsity and robustness [42].

Han Guo and Namrata Vaswani have reported the issues in video denoising through dynamic video layering. Proposed work introduced a novel layering denoising which reduces the noises and corrupted videos using sum of low rank matrix plus a sparse matrix. Authors further claimed that better performance can be achieved if the video is first decomposed into two layers, and the denoiser is applied on each layer separately [43].

## 3 Proposed Denoising Technique

Through the literature survey, it is evident that most of the techniques promising to have denoising framework are considering cases in which an image is contaminated by single noise signal. However, it is found that in practice, an image is contaminated with multiple noise type. To address this issue, a novel technique of denoising an image contaminated with multiple sources is disclosed. In this initiative, speckle noise, impulse noise, Rician noise and Gaussian noise are considered. Considering the characteristic of the noise signals Gaussian filter and impulse filter are grouped together while Rician and speckle filters are grouped together. While filtering an

image, noisy pixels contaminated with individual noise or mixed noises are processed by the filter.

To demonstrate the effectiveness of the algorithm, different noise signals are added from 10 to 60% as individual and mixed noise signal. The filtering is performed on color, synthetic and gray image.

## 4 Conclusion

In this research work, the review of various filtering techniques is presented. Different filtering techniques, algorithms and hardware approaches are proposed to restore the images contaminated with different types of noises like speckle noise, impulse noise, Rician noise and Gaussian noise. It is observed that different reported techniques are giving solutions to the individual or double noise component. However, an image can be contaminated with multiple noises. To address this issue, a novel filtering framework is proposed which considers the possibility of an image contaminated with different noises at a time. To demonstrate the robustness of the noise signal, different filters are operated in different combinations. Through the proposed technique, significant parameters like MSSIM, SSIM, MAE, execution time and PSNR can be appreciably optimized.

## References

1. Stroebel L, Zakia RD (1995) The focal encyclopedia of photography. Focal Press. p 507. ISBN 978-0-240-51417-8
2. Rohankar J (2013) Survey on various noises and techniques for denoising the color image. Int J Appl Innovation Eng Manag
3. Boyat AK, Joshi BK (2015) A review paper: noise model in digital image processing. Signal Image Process Int J (SIPIJ) 6(2)
4. Nakamura J (2017) Image sensors and signal processing for digital still cameras. ISBN 0-8493-3545-0
5. Boncelet C (2009) Image noise models. Handbook of image and video processing. Academic Press. ISBN 0-12-119792-1
6. Harikiran J, Saichandana B, Divakar B (2010) Impulse noise removal in digital images. Int J Comput Sci 10(8):39–42
7. Jayraman S, Esakkirajan S, Veerakumar T, Digital image processing. Tata McGraw Hill Education, p 272. ISBN 9781259081439
8. Rosin P, Collomosse J (2012) Image and video based artistic stylisation. Springer Publications, p 92, ISBN 9781447145196
9. Laskar RH, Bowmick B, Biswas R, Kar S (2009) Removal of impulse noise from color image. In: IEEE, region 10 conference (TENCON)
10. Maity A, Pattanaik A, Sagnika S, Pani S (2015) A comparative study on approaches to speckle noise reduction in images. In: IEEE international conference on computational intelligence and networks. https://doi.org/10.1109/cine.2015.36
11. Verma R, Ali J (2013) A comparative study of various types of image noise and efficient noise removal techniques. Int J Adv Res Comput Sci Softw Eng 3(10):618–622

12. Wang M, Zheng S (2014) A new image denoising method based on gaussian filter. In: IEEE International Conference on Information Science, Electronics and Electrical Engineering, vol 3, pp 26–28, Apr 2014
13. Rahman T, Haque MR, Rozario LJ, Uddin MS (2014) Gaussian noise reduction in digital images using a modified fuzzy filter. In: 17th IEEE Conference on Computer and Information Technology (ICCIT). https://doi.org/10.1109/iccitechn.2014.7073143
14. He K, Luan X-C, Li C-H, Liu R (2008) Gaussian noise removal of image on the local feature. In: Second international symposium on intelligent information technology application. https://doi.org/10.1109/iita.2008.552
15. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans Image Process 26(7)
16. Luisier F, Blu T, Unser M (2011) Image denoising in mixed poisson–gaussian noise. IEEE Trans Image Process 20(3)
17. Muthukumar S, Pasupathi P, Deepa S, Krishnan N (2010) An efficient color image denoising method for gaussian and impulsive noises with blur removal. In: IEEE international conference on computational intelligence and computing research. https://doi.org/10.1109/iccic.2010.5705908
18. Hara T, Guan H (2010) Color image denoising using e-neighborhood gaussian model. In: IEEE 17th international conference on image processing, 26–29 Sept 2010, Hong Kong
19. Trinh D-H, Luong M, Dibos F, Rocchisani J-M, Pham C-D, Linh-Trung N, Nguyen TQ (2014) An effective example-based learning method for denoising of medical images corrupted by heavy gaussian noise and poisson noise. In: International conference on image processing (ICIP). https://doi.org/10.1109/icip.2014.7025165
20. Mythili C, Kavitha V (2011) Efficient technique for color image noise reduction. Res Bull Jordan ACM II(III):41–44
21. Aggarwal MK, Majumdar A (2015) Mixed gaussian and impulse denoising of hyperspectral images. In: IEEE international geoscience and remote sensing symposium (IGARSS). https://doi.org/10.1109/igarss.2015.7325792
22. Kim H, Kim S (2015) Impulse-mowing anisotropic diffusion filter for image denoising. In: IEEE international conference on image processing ICIP, Jan 2015. https://doi.org/10.1109/icip.2014.7025591
23. Iqbal N, Ahmad K, Shahjehan W (2017) High density impulse noise reduction by denoising neighbor pixels. In: 13th International conference on emerging technologies (ICET). https://doi.org/10.1109/icet.2017.8281722
24. Xiong B, Yin Z (2012) A universal denoising framework with a new impulse detector and nonlocal means. IEEE Trans Image Process 21(4)
25. Tariyal S, Aggarwal HK, Majumdar A (2015) Hyperspectral impulse denoising with sparse and low-rank penalties. In: 7th Workshop on hyperspectral image and signal processing: evolution in remote sensing (WHISPERS), https://doi.org/10.1109/whispers.2015.8075397
26. Newlin DR, Christopher CS (2015) Random valued impulse denoising using adaptive weighted interpolation. In: International conference on control, instrumentation, communication and computational technologies (lCCICCT), Dec 2015
27. Indu S, Ramesh C (2009) Image fusion algorithm for impulse noise reduction. In: International conference on advances in recent technologies in communication and computing, Oct 2009
28. Majumdar A, Ansari N, Aggarwal HK (2015) Hyper-spectral impulse denoising: a row-sparse blind compressed sensing formulation. In: IEEE International conference on acoustics, speech and signal processing (ICASSP), Apr 2015
29. Gupta M, Garg A (2017) An efficient technique for speckle noise reduction in ultrasound images. In: 4th International conference on signal processing and integrated networks (SPIN), Feb 2017
30. Parthiban L, Subramanian R (2006) Speckle noise removal using contourlets. In: International conference on information and automation, Dec 2006
31. Garg A, Khandelwal V (2017) Speckle noise reduction in medical ultrasound images using coefficient of dispersion. In: International conference on signal processing and communication (ICSC), July 2017

32. Barcelos C, Vieira L (2014) Ultrasound speckle noise reduction via an adaptive edge-controlled variational method. In: IEEE international conference on system, man and cybernetics, Oct 2014
33. Rahman M, Kumar M, Borucki B, Nowinski KS (2013) Speckle noise reduction of ultrasound images using extra-energy reduction function. In: International conference on informatics, electronics and vision (ICIEV), May 2013
34. Murakami T, Ogawa K (2018) Speckle noise reduction of optical coherence tomography images with a wavelet transform. In: 14th IEEE international colloquium on signal processing & its applications (CSPA), Mar 2018
35. Udomhunsakul S, Wongsita P (2004) Ultrasonic speckle denoising using the combination of wavelet transform and wiener filter. In: The Second asian and pacific rim symposium on biophotonics, Dec 2004
36. MaluGan R, Terebeu R, Germain C, Bordal M, Ciulariu M (2015) Speckle noise removal in ultrasound images using sparse code shrinkage. In: The 5th IEEE international conference on e-health and bioengineering—EHB
37. Barcelos C, Vieira L (2014) Ultrasound speckle noise reduction via an adaptive edge-controlled variational method. In: IEEE international conference on system, man and cybernetics
38. Misra D, Sarker S, Dhabal S, Ganguly A (2013) Effect of using genetic algorithm to denoise MRI images corrupted with rician noise. In: IEEE international conference on emerging trends in computing, communication and nanotechnology (ICECCN 2013). https://doi.org/10.1109/ice-ccn.2013.6528481
39. Nowak RD (1999) Wavelet-based rician noise removalfor magnetic resonance imaging. IEEE Trans Image Process 8(10)
40. Chang H-H, Hsieh T-J, Ting Y-N, Chu W-C (2011) Rician noise removal in mr images using an adaptive trilateral filter. In: 4th International conference on biomedical engineering and informatics (BMEI). https://doi.org/10.1109/bmei.2011.6098281
41. Liu X, Dai T (2018) A new guided image denoising by principal component analysis with local pixel grouping. In: Fourth IEEE international conference on multimedia big data (BigMM)
42. Bochinski E, Jongebloed R, Tok M, Sikora T (2018) Regularized gradient descent training of steered mixture of experts for sparse image representation. In: 25th IEEE international conference on image processing (ICIP). https://doi.org/10.1109/icip.2018.8451823
43. Guo H, Vaswani N (2018) Video denoising via dynamic video layering. IEEE Signal Process Lett 25(7)
44. Sotiriou E, Xydis S, Siozios K, Economakos G, Soudris D (2015) Hardware accelerated rician denoise algorithm for high performance magnetic resonance imaging. In: 4th International conference on wireless mobile communication and healthcare—transforming healthcare through innovations in mobile and wireless technologies (MOBIHEALTH). https://doi.org/10.1109/mobihealth.2014.7015951, January 2015

# Algorithmic Trading Using Machine Learning and Neural Network


**Devansh Agarwal, Richa Sheth, and Narendra Shekokar**

**Abstract** Machine learning models are becoming progressively predominant in the algorithmic trading paradigm. It is known that a helpful data is taking cover behind the noisy and enormous information that can give us better understanding on the capital markets. There are multiple issues, which are prevalent as of now by including the overfitting model, irrelevant/noisy data used for training models due to which the efficiency of the existing models fail. In addition to these problems, the existing authors are facing issues with the dispersion of daily data, poor presentation, and the problems faced with too much or too little data information. The main objective in this undertaking is to discover a technique that choose gainful stocks ordinarily by mining the public information. To accomplish this, various models are assembled to foresee the everyday return of a stock from a lot of features. These features are built to be more dependent on the cited and outside information that is accessible before the forecast date. Numerous sorts of calculations are utilized for anticipating/forecasting. When considering machine learning, regression model is implemented. Neural networks, as a wise information mining strategy and profound learning methods are valuable in learning complex types of information by utilizing the models of regulated learning, where it progressively learns through datasets and experience. Because of high volumes of information produced in capital markets, machines would master different designs, which in turn makes sensibly great predictions. In the current proposed venture, LSTM model is utilized through RMS prop optimization for anticipating future stock estimations. Also, feed forward multi-layer perceptron (MLP) is utilized along with recurrent network to foresee an organization's stock worth by depending on its stock share price history. The cycle in the

D. Agarwal (✉) · R. Sheth · N. Shekokar
Dwarkadas J. Sanghvi College of Engineering, Mumbai Univeristy, Mumbai, India
e-mail: devanshagarwal483@gmail.com

R. Sheth
e-mail: richasheth46@gmail.com

N. Shekokar
e-mail: narendra.shekokar@djsce.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_33

financial exchange is clearly with a ton of vulnerability, so it is profoundly influenced by a great deal of numerous elements. Nonetheless, the outcomes acquired show that the neural networks will outperform the existing one-dimensional models.

**Keywords** Algorithmic trading · Machine learning · Neural networks · Multi-layer perceptron (MLP) · Recurrent network · Overfitting models · LSTM

## 1 Introduction

A trader can "play" the stock exchange by utilizing the forecast models over an assessment period. The financial specialists will utilize a technique educated by the given model with which they would then be able to compare the straightforward methodology of purchasing and holding the stock over a whole period. Development of the stock market can likewise be anticipated by technical analysis [1]. The two fundamental parts of technical analysis are cost and volume, and based on that two information, the entire securities exchange can be anticipated. Stock market development is only a blend and match of mathematics and human brain science, and technical analysis is mainly about these two attributes. When the market moves positive individuals contribute expecting a further sure development, however, when little downtrend is seen individuals book benefits expecting, little produced benefits would be lost. At the point when the market goes down, individuals hold feeling that they would exit if the market falls further. They normalize it out and accordingly increment their misfortunes. The right route is to cut the misfortunes by putting right plug misfortunes and include more stocks with the benefits while keeping up following stop misfortunes.

Trading requires a ton of consideration and sensitivity to the market. Experienced brokers depend on numerous wellsprings of data, for example, news, verifiable information, yearly reports, and insiders. Risk is high and many are factors should have been thought of [2]. Hence, some financial organizations depend simply on machines to make trades. Stock value developments are to some degree monotonous in nature in the time arrangement of stock prices. The expectation highlights of the framework attempt to foresee the stock return in the time arrangement esteem via preparing the model or investigating the graphs pattern of specialized indicators, which includes delivering a yield and rectifying the mistake that implies a machine with rapid web associations can execute many trades during a day making a profit from a little distinction in prices. Machine learning in account has become more conspicuous as of late because of the accessibility of huge measures of information and more moderate computing power. It challenges existing acts of displaying and model use and drives an interest for pragmatic answers for how to deal with the intricacy relating to these techniques. Machine learning includes giving data samples to an algorithm, normally obtained from previously valid samples. The information tests consist of factors called indicators, just as target variables, which is a normal outcome. The calculation utilizes the indicator factors to foresee the objective variables. Leading

banks and other organizations are sending AI to smooth out their cycles, advance portfolios, decline risks, and guarantee advances in addition to other things.

Neural networks have been utilized progressively in various business applications, including anticipating and promoting research arrangements. In certain zones, for example, risk prediction or fraud detection, they are the unquestionable pioneers. The significant fields in which neural networks have discovered application are monetary activities, venture planning, trading, business analysis, and maintenance of products. Neural networks can be applied beneficially by a wide range of traders, so in case you are a trader and you have not yet been acquainted with neural network, this paper will take you through this technique for specialized investigation and tell you the best way to apply it to your trading style. Neural networks are cutting edge in software engineering. They are basically trainable algorithms that attempt to imitate certain parts of the working of the human mind. This gives them a remarkable, self-training capacity, the capacity to formalize unclassified data, and, above all, the capacity to make gauges dependent on the authentic data they have at their disposal. A significant confusion is that neural networks for an anticipating tool that can offer counsel on the proper behavior in a specific market circumstance. Neural networks do not make any estimates. Rather, they investigate value information and reveal openings. Utilizing a neural network, you can settle on an exchange choice dependent on completely analyzed information, which is not really the situation when utilizing customary specialized investigation strategies. For a genuine, thinking trader, neural networks are a cutting-edge device with incredible potential that can identify unobtrusive non-direct interdependencies and examples that different techniques for specialized investigation cannot reveal.

Conventional strategies, for example, logistic regression and classification, are based on linear models, while neural networks are self-changing techniques dependent on training information, so they can tackle the issue with a little information about its model and without compelling the forecast model by including any additional assumptions. Also, neural networks can discover the connection between the information and yield of the framework regardless of whether this relationship may be extremely convoluted on the grounds that they are general function approximators.

One of the major problems faced while implementing the machine learning algorithms was the overfitting problem. Overfitting refers to a model that models the training data too well. It happens when a model learns the detail and noise within the training data to the extent that it negatively impacts the performance of the model on new data [3]. This suggests that the noise or random fluctuations within the training data are picked up and learned as concepts by the model. The matter is that these concepts do not apply to new data and negatively impact the model's ability to generalize. Overfitting is more likely with nonparametric and nonlinear models that have more flexibility when learning a target function. This problem can be considerably solved using the classical and modern deep learning approach [4]. The principal objective of the venture was to consider and apply neural networks that are more productive in anticipating the cost of a stock when contrasted with many machine learning algorithms [5]. An attempt to use six learning algorithms on tata consultancy services dataset, three machine learning models consisting of support vector machine

(SVM) and linear, logistic regression versus two deep learning models including multi-layer perceptron (MLP), Elam recurrent network and long short-term memory (LSTM) is used. The paper has proposed a prescient model based on neural networks for foreseeing stock market changes. Tests show that the prediction error of this model is around 1.5%. The structure of the paper is organized thusly. Related documents and studies are presented in Sect. 2. The process of building the dataset and dataset evaluation is described in Sect. 3. In Sect. 4, it is shown how to apply linear, logistic regression and SVM, MLP, recurrent network, and LSTM for this task. The experimental stock prediction results are analyzed in Sect. 5. Conclusion and future work are deduced in Sect. 6.

## 2 Related Works

| Paper name | Author | Advantages | Limitations | Inference |
|---|---|---|---|---|
| Introduction to financial forecasting [6] | 1. Yaser S. Abu-Mostafa<br>2. Ammir F. Atiya | The linear models act well. The linear classification model and linear regression give comparative outcomes here. The regression approach and classification approach both can catch the basic differences | The supporting vector machines do not act alright. One potential explanation is that for these incredibly loud information, direct basic models can act better. SVM classifier's presentation is particularly awful | It exhibits the utilization of hints calculation for foreign trade exchanging over a time of 32 months<br>It additionally talks about the fundamental methodology for gauging information and examines the loud idea of financial information |

(continued)

(continued)

| Paper name | Author | Advantages | Limitations | Inference |
|---|---|---|---|---|
| Predicting stock market returns with machine learning [7] | 1. Alberto G. Rossi | Boosted tree regression outperforms those created by benchmark models as far as both mean squared mistake and directional precision. They additionally create gainful portfolio assignments for mean-change financial specialists even at the point when market erosions are represented | Number of questions with respect to their precision as the greater part of the regressors considered are extremely diligent, making measurable deduction not exactly clear. Data snooping around might be a wellspring of concern if specialists are trying for a wide range of model particulars and report just the factually critical ones | A semi-parametric technique known as boosted regression trees (BRT) is used to gauge stock returns and unpredictability at then month to month recurrence based on enormous arrangements of molding data without forcing solid parametric presumptions, for example, linearity or monotonicity. Results recommend that the connection between indicator factors and the ideal portfolio allotment to unsafe resources is exceptionally non-straight |

(continued)

(continued)

| Paper name | Author | Advantages | Limitations | Inference |
|---|---|---|---|---|
| Predicting stock prices using data mining techniques [8] | 1. Qasem A Al Radaideh<br>2. Adel Abu Assaf<br>3. Eman Alnagi | The resultant classification exactness from the decision tree model is not exceptionally high for the preparation information utilized, and it fluctuates from one organization to another. Organization's presentation in the financial exchange is influenced by news about the organization, money related reports, the general presentation of the market, political functions, and political choices | The resultant classification exactness from the decision tree model is not exceptionally high for the preparation information utilized, and it fluctuates from one organization to another. Organization's presentation in the financial exchange is influenced by news about the organization, money related reports, the general presentation of the market, political functions, and political choices | To choose the better planning for purchasing or selling stocks dependent on the information extricated from the authentic costs of such stocks, choice is taken dependent on decision tree classifier . The CRISP-DM strategy is utilized over genuine authentic information of significant organizations |
| Neural networks, financial trading, and efficient markets hypothesis [9] | 1. Andrew Skabar<br>2. Ian Choete | Free from data snooping. Incorporated exchanging cost of 0.1% per exchange, which is at present the estimated commission for web-based exchanging | These boot-strapped tests contain similar dispersion of day by day returns as the first arrangement, however, do not have any sequential reliance present in the first. | Centrality of the profits accomplished is finished utilizing bootstrapping strategy An exchanging technique reliant on recorded value information can be utilized to accomplish returns in a way that is better than those accomplished utilizing a purchase and hold procedure. |

(continued)

(continued)

| Paper name | Author | Advantages | Limitations | Inference |
|---|---|---|---|---|
| Neural network forecasts of Canadian stock returns using accounting ratios. [10] | 1. Dennis Olson<br>2. Charles Mossman | Neural network outperforms relating conventional regression strategies. It selects only most noteworthy, positioned stocks, or those with the most elevated likelihood of having returns | There are drawbacks from introducing either an excess of data (OLS assessment models) or too little information (binary logit models) | The backpropagation neural network, which considers non-direct connections among information and yield factors, outperforms the best regression options for both point assessment and in ordering firms expected to have either high or low returns |
| Artificial neural network approach for stock price and trend prediction [11] | 1. Nasimul Hasan<br>2. Risul Islam Rasel | Two assessment measures, mean average percentage error (MAPE) and root mean square error (RMSE) and proposed models—1 day ahead, 5 days ahead, and 10 days ahead are utilized in this investigation to expand precision and limit mistake | Just windowing administrator was utilized in this research for information preprocessing step, and the examination was planned dependent on just the New York stock exchange | Artificial neural network (ANN) is utilized alongside windowing administrator which is profoundly proficient for working with time arrangement information for anticipating securities exchange cost and pattern. With his technique, it can create a reasonable outcome with a small error |

## 3 Methodology

In the setup, the anticipation of the pattern of the stock in a particular day is attempted. Also, one can foresee the specific return of a stock in regression setup. For straightforwardness, the first attempt is using direct models as follows: classification model using the logistic regression and the regression model using the linear regression.

At that point, actualize SVM models (classifier and regression) for implementing conceivable nonlinear conventions which use kernel. Anyway, used model is dynamic instead of fixed, contingent upon the date at which a return is to be anticipated.

Analysis of regression starts with an historical dataset in which the objective qualities are known to create the valuation work. After the valuation capacities are resolved, they would then be able to be applied to the dataset as a piece of the forecast. Linear regression aims at modeling a relationship between two variables to predict a pattern. In this model, the TCS dataset is loaded and picked the target variable. Once the target variable is picked, training and validation process is performed. A specific ratio is used for this process. After this, the linear regression algorithm is used to predict the stock price. This is confirmed using the validation close and the prediction close values.

Logistic regression is a strategy for making predictions when the needy variable is polarity, and the independent factors are ceaseless and additionally discrete. The upside of logistic regression is that through the expansion of a suitable connection capacity to the typical direct relapse model, the factors may either be persistent or discrete or any mix of the two kinds, and they do not really have ordinary distributions. The stock qualities are then anticipated utilizing the produced model from the training dataset. The yield is given as a likelihood score which has range from 0 to 1. If the anticipated likelihood score is more than 0.5 in this investigation, at that point, those stocks are considered to have a rising pattern and make a benefit in future [12]. Therefore, as a result, if the stock has a high probability, go long on it.

The SVM utilizes a gaussian part and was improved over sigma and the edge classifier utilizing cross validation. The preparation of the SVM requires three boundaries: C-cost of misclassification, lambda, and a kernel [1]. When the calculation cycle begins, the information is taken care of to the model. The assessment of this model returns two vectors: the determined appropriate activity for each assessed day and the level of having a place for every one of the dissected classes (being just two: a buy or a deal). Joining these two yields restores a vector of qualities from the reach $[-1, 1]$ that assigns the chosen request for each testing day. This vector is advanced utilizing the improvement module. In the optimization cycle, it is observable that the SVM will in general have better outcomes in anticipating the down moving periods, anyway the wonders get returned in the up moving one [13].

Multi-layer perceptron neural networks are the Levenberg–Marquardt error back propagation, wherein the organization learns the variation in the informational collection and uses gradient as well as Jacobean for performance measure to legitimize loads of the networks the converse way concerning the slope vector of error work which is generally a regularized entirety of squared error [14]. It uses gradient as well as Jacobean for performance measure.

For anticipating the direction of changes of the qualities in the following day, none of these techniques are superior to the basic straight regression model. But the error of the forecast of the measure of significant worth changes utilizing MLP neural organization is less than linear regression strategy [14]. Also, when the feed forward MLP neural organization predicts the heading of the progressions effectively, the measure of progress is totally near the genuine one.

Recurrent neural networks (RNNs) unlike the multi-layer perceptron (MLP) take contributions from two kinds of sources, one is from the present and the other is from the past [1]. To prepare Elman organization, the error backpropagation with momentum and adaptive learning rate is utilized. The loads of the organization in this calculation are changed by slope (with momentum), past change in the organization loads, and learning rate [14]. Data obtained from such sources is to be utilized for displaying networks response to the fresher arrangements of input information. This is conceivable with criticism loops wherein yield at any case can be taken as a contribution to its resulting example. This implies RNN needs memory. Each contribution with its gigantic heaps of information should be put away in certain layers of RNNs. This stored information or data is recursively utilized in the organization as it will be swept forward for managing more current models.

Long short-term memory (LSTM) is just a model of recurrent neural networks; these networks are more grounded in recognizing just as learning in longer term conditions. Ordinary RNN includes a straightforward organization alongside feedback circles, while LSTM comprises memory squares or cells instead of a solitary organization layer. Each cell comprises three gates just as a cell state machine which controls information moves through such cells [1]. The LSTM model can be tuned for different parameters, for example, changing the quantity of LSTM layers, adding dropout esteem, or expanding the quantity of epochs to further identify whether the stock price will go up or down.

## 4 Proposed Model

In this paper, two sorts of strategies, linear regression and LSTM recurrent network, are utilized to foresee an organization's stock price dependent on its stock share value history. The test results show that the use of LSTM recurrent network is more encouraging in anticipating stock worth changes instead of linear regression technique.

The model which is proposed has a block diagram as shown in Fig. 1.

Figure 1 comprises of stacking a dataset, choosing an objective variable, preparing and approving the dataset, applying the Linear regression/LSTM recurrent network, and subsequently making the stock forecast.

Stacking the dataset is the beginning measure for this cycle. For this undertaking, tata consultancy services (TCS) dataset is utilized, considering its stock cost from November 2013 to March 2020. The dataset comprises of the accompanying credits: date, open, high, low, close, adjusted close, volume, and the normal of high and low cost of the day.

The open and close segment shows the beginning and the consummation estimation on that specific day. High and low section represents the most noteworthy and the least noteworthy which the stock has reached on that specific day. Volume segment speaks to the quantity of shares exchanged.

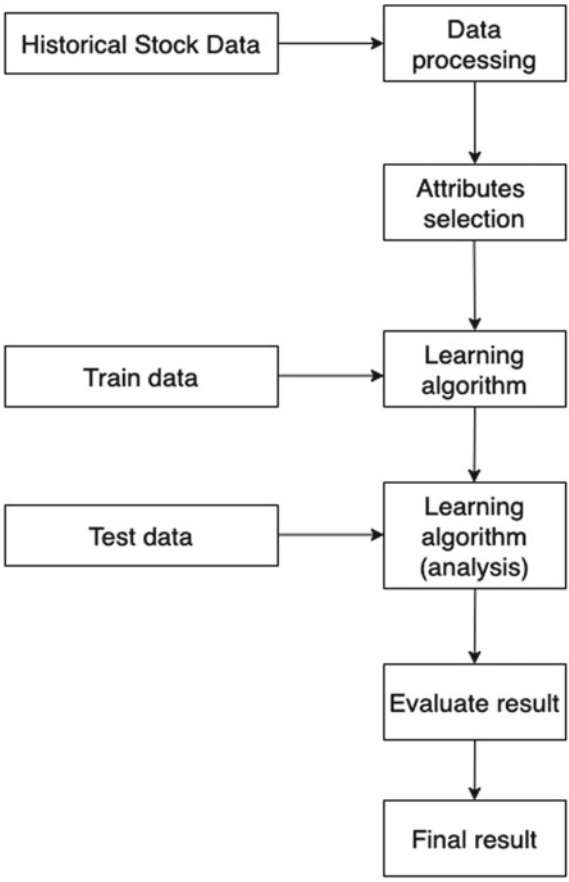



**Fig. 1** Flow of proposed model

The subsequent stage is choosing an objective variable. Since all the benefit and damage count is normally founded on the shut value, close is picked as objective variable. Once the dataset is stacking and the objective variable is fixed, the dataset is prepared and approved.

The regression model returns a condition that decides the association between the independent factors and furthermore the variable quantity. The condition for regression is composed as:

$$Y = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

In the equation above,$x_1, x_2, \ldots, x_n$ represent the independent variables while the coefficients, $\theta_1, \theta_2, \ldots, \theta_n$ represent the weights. The dataset is arranged in an ascending order and afterward made a different dataset with the goal that any new component does not get influenced by the main information. After this, the linear

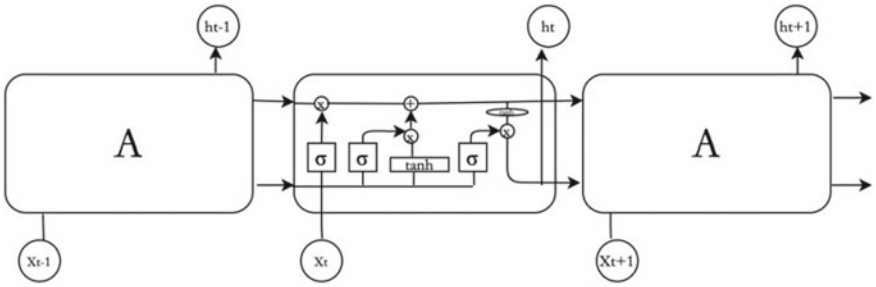



**Fig. 2** Illustration of LSTM model

regression model is utilized, and forecasts are made. Presently, the subsequent model is the neural organization LSTM model for foreseeing stock cost.

Figure 2 shows the basic architecture of LSTM model which has been used in this project. Our justification is because LSTM can store past information that is remarkable and ignore the information that is not. LSTM can add or take out information to the cell state, obliged by structures called gates. Gates are used for, then again, letting information through. Passages are made of a sigmoid neural net layer and a point-wise duplication movement.

This model, LSTM has three entryways:

- The input gate: The information entryway adds data to the cell state.
- The forget gate: It wipes out the information that is not, now required by the model.
- The output gate: At LSTM, it chooses the data to be appeared as yield.

Sequential is utilized for presenting the neural organization, and this model incorporates the LSTM layer, dropout for thwarting overfitting with dropout layers, and dense to incorporate an associated related neural association layer.

For information normalization to scale the readiness dataset, estimator scale is used with numbers some place in the scope of zero and one. The LSTM layer is incorporated with the going with contentions: 50 units are the dimensionality of the yield space, whether to return the last yield in the sequence is fundamental for stacking LSTM layers, so the following LSTM layer has a three-dimensional arrangement input, and information shape is the condition of the planning dataset. To mastermind the model, the Adam optimizer is used for learning parameters and set the loss as the mean squared error. From that point forward, the model is fitted to run for 20 epochs (the ages are the events the learning count will work through the entire planning set) with a batch size of 32.

Prior to predicting future stock costs, the test set is adjusted by solidify the train set and the test set on the 0 pivot, set 60 as the time step again, use estimator scale, and reshape data. By then, inverse transform places the stock expenses in a regular significant affiliation.

Henceforth, this model is utilized to anticipate stock cost and furthermore can be tuned for different limits, for instance, changing the amount of LSTM layers, including dropout regard, or expanding the quantity of epochs.

## 5 Experimentation

In this project, two algorithmic models were implemented, LSTM and linear regression.

**Algorithm 1** Pseudo code for algorithmic Trading using LSTM model

Input: x_train, y_train, x_test, y_test

Output: Accuracy

1. Procedure: LSTM MODEL (x_train, y_train)
2. Parameters: batchSize = 32; num_epochs = 20; verbose = 2
3. Model = Sequential ()
4. Two hidden layer Stacked LSTM-Regardless of whether to restore the last yield in the output sequence or the full sequence
   model.add(LSTM(100, return_sequences = True))
   model.add(LSTM(units = 50))
5. Dense Layers-This layer is connected to the LSTM layers of the model and is often used to output the prediction.
   model.add(Dense(1))
6. Compile Function-It is an effectiveness step. The optimizer calculation is used to prepare the organization and the loss work is used to assess the organization that is limited by the optimization calculation.
   model.compile(loss = 'mean_squared_error', optimizer = 'adam')
7. Fitting a model
8. Model Evaluation
9. End Procedure

Above is the pseudo-code for the algorithm used. The environment on which the code was run is google colab (jupyter notebook environment) in addition to the libraries sklearn and keras.

## 6 Result Analysis

Here, 70/30 ratio for splitting the data into training and validation set is used, that is, 70% of the data will be used to train the model, and 30% will be used to validate. More training data is essential since it implies the model sees more models and accordingly

ideally finds a superior solution. This method is adopted since the dataset used is enormous, and it covers the data for the past 7 years.

In Figs. 3 and 4, the orange line depicts the close values for the training dataset, the blue line shows the validating close values, and the yellow line depicts the predicted close value for the same validating close dataset. Linear regression is a basic method and very simple to decipher, yet there are a couple of clear hindrances. One issue in utilizing regression calculations is that the model overfits to the date and month segment. It can clearly be seen that the predicted close value for the LSTM model is much more accurate than that of the linear regression model. This can further be confirmed using the rms values for each.

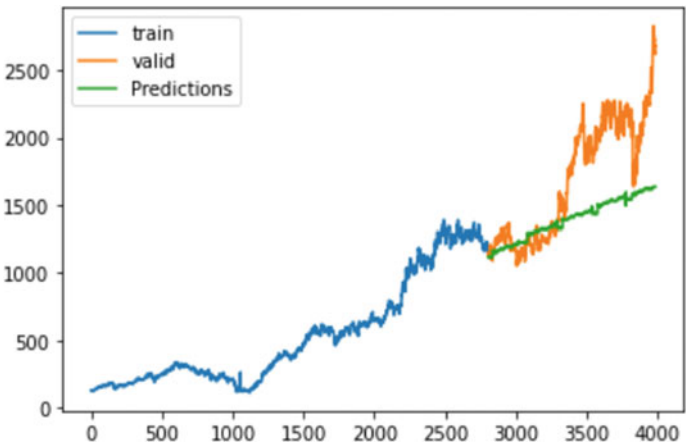



**Fig. 3** Regression model prediction

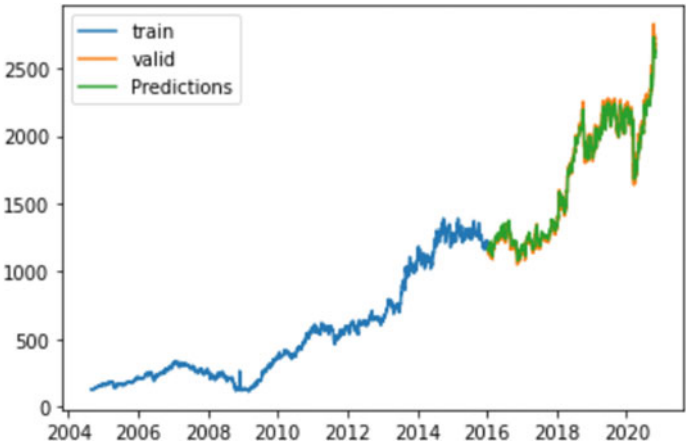



**Fig. 4** LSTM model prediction

```
preds = model.predict(x_valid)
rms=np.sqrt(np.mean(np.power((np.array(y_valid)-np.array(preds)),2)))
rms
```

```
426.4024276628599
```

**Fig. 5** RMS linear regression

```
rms=np.sqrt(np.mean(np.power((valid-closing_price),2)))
rms
```

```
48.66518525086109
```

**Fig. 6** RMS LSTM model

Figures 5 and 6 show the root mean square of both the models. First one being the linear regression model and the second one being the LSTM model. The RMS value of the linear regression model is 426 in comparison to the 48 RMS of the LSTM model. Since, the RMS value of the LSTM model is much lesser than that of linear regression model, once again it is proved that the neural networks model is much more accurate than the linear regression model.

## 7 Conclusion

The proposed model considers two algorithms, which can be used for algorithmic trading, liner regression, and long short-term memory (LSTM). It has been successfully shown that the LSTM model gives promising results in comparison with the linear regression model. The major merit of this work is that major speculations in the market can be curbed down by using these algorithms. Since, the proposed model uses a machine to learn from the past data and hence it predicts the stock prices, where it can do a better job at it, as compared to humans.

## 8 Future Scope

Later on, an attempt to discover strategies that give more steady forecasts can be performed. The proposed research work intends to apply other recently proposed techniques of machine learning and deep learning models, which are more current in the field of AI research and professed to have great speculation capacity because of the use of enormous edge ideas. From the point of view of machine learning, it strives at blending various models on the train models with more information

consistently. From a neural network perspective, different models can be coordinated on the equivalent dataset. Moreover, subsequent to getting the outcomes from every implemented model, it may be well-demonstrated with which models are remaining as the most productive. Further, after the thorough comparison of machine learning and neural network models, it is desired to segregate and come up with sector-wise algorithms.

## References

1. Dhenuvakonda P, Anandan R, Kumar N (2020) Stock price prediction using artificial neural networks. J Crit Rev
2. Chowdhury C (2018) Stock market prediction for algorithmic trading using machine learning techniques & predictive analytics
3. Shah VH (2007) Machine Learning techniques for stock prediction
4. Rice L, Wong E, Kolter JZ (2020) Overfitting in adversarial robust deep learning. In: Proceedings of the 37th international conference on machine learning
5. Rose LK (2018) Automated stock market trading using machine learning
6. Rossi AG (2019) Predicting stock market returns with machine learning
7. Al-Radaideh QA, Assa AA, Alnagi E (2019) Predicting stock prices using data mining techniques. In: 2nd International conference on advances in science and technology (ICAST)
8. Skabar A, Choete I (2002) Neural networks, financial trading, and efficient markets hypothesis
9. Olson D, Mossma C (2003) Neural network forecasts of Canadian stock returns using accounting ratios February 2003. Int J Forecast
10. Hasan N, Rasel RI (2016) Artificial neural network approach for stock price and trend prediction. In: International conference on advanced information and communication technology
11. Abu-Mostafa YS, Atiya AF (1996) Introduction to financial forecasting. Kulwar academic publishers, Netherlands
12. Hargreaves C, Dixit P, Solanki A (2013) Stock portfolio selection using data mining approach. IOSR J Eng
13. Szklarz J, Rosillo R, Alvarez N, Fernández I, Garcia N (2018) Application of support vector machine on algorithmic trading. In: International conference artificial intelligence (ICAI)
14. Naeini MP, Taremian H, Hashemi HB (2010) Stock market value prediction using neural networks. In: 2010 international conference on computer information systems and industrial management applications (CISIM)

# Analysis on Intrusion Detection System Using Machine Learning Techniques


**B. Ida Seraphim and E. Poovammal**

**Abstract** Intrusion detection system [IDS] is a significant base for the network defence. A huge amount of data is generated with the latest technologies like cloud computing and social media networks. As the data generation keeps increasing, there are chances that different forms of intrusion attacks are also possible. This paper mainly focuses on the machine learning (ML) techniques for cyber security in support of intrusion detection. It uses three different algorithms, namely Naïve Bayes classifier, Hoeffding tree classifier and ensemble classifier. The study is performed on emerging methods and is compared with streaming and non-streaming environment. The discussion on using the emerging methods and challenges is presented in this paper with the well-known NSL_KDD datasets. The concept of drift is induced in the static stream by using the SEA generator. Finally, it is found that the ensemble classifier is more suitable for both the environments with and without concept drift.

**Keywords** Intrusion detection · Cyber security · Data mining · Machine learning · Naïve bayes · Hoeffding tree · Ensemble classifier

## 1 Introduction

With the tremendous increase in the usage of Internet, network security is remaining as the need of the hour. The traditional security techniques like firewall, user authentication and data encryption will provide the first security defence line [1]. This first line of security will not be enough to provide the required security. The second line of security is recommended and that can be provided by the intrusion detection system (IDS). The combination of these two lines of security enhances the overall network security.

B. I. Seraphim (✉) · E. Poovammal
Department of Computer Science and Engineering, SRMIST, Chennai, India
e-mail: idaserab@srmist.edu.in

E. Poovammal
e-mail: poovamme@srmist.edu.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_34

An intrusion detection system (IDS) is a device or software that is used to detect or monitor the existence of an intruder attempting to breach the network or a system [4]. The security and the communication of digital information in a secure manner are more important due to the tremendous growth and usage of the Internet. The major issue of the industry, government and commerce worldwide is the intrusion. An IDS that provides a fast and precise alert to the intrusion is the need of the hour.

Security information and event management (SIEM) system centrally collects and alerts the network administrator about the malicious activities or policy violations. An IDS monitors the network or system for the intrusion. The output is combined using SIEM from multiple sources, and the false alarm filtering method is used to differentiate between malicious attack and false alarm.

Intrusion detection system scope ranges from simple antivirus software to hierarchical systems that monitor the entire network's traffic. An IDS is classified into two types based on architecture as a host-based intrusion detection system (HIDS) or network-based intrusion detection system (NIDS) [4]. HIDS identifies intrusion by analysing system calls, application logs and file system modifications. NIDS identifies intrusion by analysing and examining network traffic, network protocols and network data for suspicious activities. Based on the detection mechanism, an IDS can be classified into two types as anomaly-based IDS and signature-based IDS. Signature-based detection identifies the attack by observing the patterns, signatures and comparing those patterns and signatures with the already stored signatures. A signature is a pattern that describes the attack or exploits that are well known [3]. The new signatures need to be updated regularly in the databases. The signature-based detection technique is not suitable for novel attacks.

Anomaly-based detection identifies malicious events by finding the deviation of the monitored traffic from the normal profile. The previously unseen (legitimate) system behaviours may be considered anomalies because there is an increase in the false alarm rate (FAR). This is the main weakness of the anomaly-based detection technique. Three main limitations contribute to network security challenges [2]. 1. The volume of the network data is huge. 2. Monitoring must be in-depth to improve efficacy and accuracy. 3. The diversity of data navigating through the network, and the number of different protocols is used. Figure 1 shows the basic workflow carried on to detect the intrusion in the data.

The NSL_KDD dataset is taken for the analysis. It is the benchmark dataset taken from Canadian institute of cyber security. The dataset has 41 attributes, and the final attribute is the class attribute that shows whether the instance is normal or anomaly. The dataset is uploaded to the massive online analysis framework, and the whole dataset is split into test and train data. Then, the chosen machine learning techniques are applied in non-steaming and streaming environment. In streaming environment, the algorithms are applied to stream with and without concept drift. Finally, the performance measure like accuracy, kappa and time is evaluated.

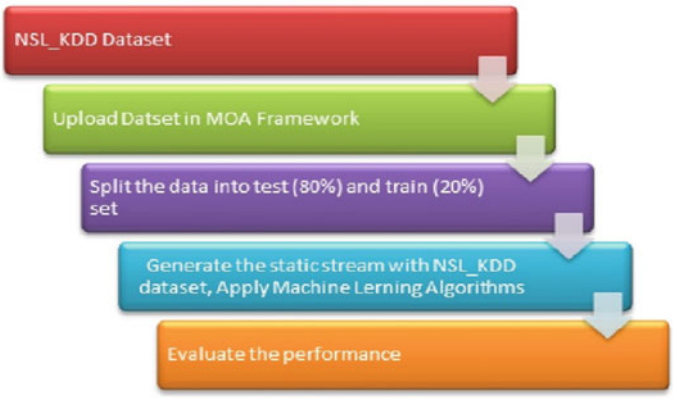



**Fig. 1** Basic workflow diagram

## 2 Related Work

The main aim of an intrusion detection system (IDS) is to find the abnormal behaviours inside and outside the network.

| Authors | Title of the paper | Algorithms used | Dataset | Remarks |
|---|---|---|---|---|
| Asmah Muallem et al. (Jounal of Information Security, 2017) | Hoeffding Tree Algorithms for Anomaly Detection in Streaming Datasets: A Survey | Hoeffding tree Restricted Hoeffding tree Accuracy Updated Ensemble | KDD Cup'99 NSL-KDD | Handling anomalies in stream environment [10] |
| Prof. Dr. P. K. Srimani et al. (Wseas Transactions on Computers, 2016) | Mining Data Streams with Concept Drift in Massive Online Analysis Framework | Naïve Bayes classifier | Synthetic dataset generator algorithms Benchmark dataset electricity, airline and forest cover dataset. | Naïve Bayes gives less accuracy. Kappa values are less, and time consumption is more [16] |

(continued)

(continued)

| Authors | Title of the paper | Algorithms used | Dataset | Remarks |
|---|---|---|---|---|
| Preeti Mishra et al. (IEEE Communication Surveys & Tutorials, 2018) | A Detailed Investigation and Analysis of using Machine Learning Techniques for Intrusion Detection | Decision Tree<br>ANN<br>Naïve Bayes Classifier<br>SVM<br>Genetic Algorithm. | NSL-KDD | Challenges<br>1. Requires lot of data for training and classification<br>2. Complexity in training huge amount of data<br>3. High performance hardware required for training [4] |
| Yasir Hamid et al. (International Journal of Network Security, 2018) | Benchmark Datasets for Network Intrusion Detection: A Review | K-NN Classifier | Full KDD99, Corrected KDD, NSL-KDD, 10% KDD, UNSW, Caida, ADFA Windows, UNM Dataset | Ensemble algorithms can still improve the accuracy [17] |
| Ketan Sanjay Desale et al. (IEEE International Conference on Computing Communication Control and Automation, 2015) | Efficient Intrusion Detection System using Stream Data Mining Classification Technique | Accuracy weighted Ensemble<br>Naïve Bayes<br>Hoeffding Tree<br>Accuracy Updated Ensemble | NSL-KDD | 1. Hoeffding Tree gives less accuracy in less time.<br>2. Accuracy Weighted Ensemble gave high accuracy but took more time [8] |
| Ngoc Tu Pham et al. (ACMW, Brisbane, Australia, 2018) | Improving Performance of Intrusion Detection System using Ensemble Methods and Feature Selection | Bagging and Boosting combined with tree-based algorithms (such as J48, Random Forest, Random Tree, REPTree | NSL-KDD | 1. Single dataset was used to evaluate the classifiers.<br>2. Accuracy can be improved further [7] |
| Lidong Wang (Journal of Computer Networks 2017) | Big Data in Intrusion Detection System and Intrusion Prevision System | Reviewed different data mining and machine learning techniques | KDD Cup'99 | Biggest challenge is the Real time monitoring [18] |

(continued)

(continued)

| Authors | Title of the paper | Algorithms used | Dataset | Remarks |
|---|---|---|---|---|
| Kai Peng et al. (Special Section on Cyber Physical Social Computing and Networking, IEEE Access, 2018) | Clustering Approach Based on Mini Batch K-means for Intrusion Detection System Over Big Data | K-Means<br>Mini Batch K-means | KDD Cup'99 | 1. K-means cannot effectively determine which K value is best.<br>2. PMBKM – effective when both K value is small and big [6] |

## 3 Machine Learning (ML) Techniques

There are two different categories of machine learning (ML) techniques, namely supervised and unsupervised learning. Supervised learning is that in which labelled instances are used for training the model. Supervised learning uses an input variable ($x_1$) and an output variable ($y_1$), and it uses a classification algorithm to learn the mapping function between input and output [9]. Equation 1 shows the mapping of the input and output function.

$$y_1 = f(x_1) \tag{1}$$

Supervised learning problems are further classified into classification and regression problems. Classification models are those in which the class label is known. The classification model tries to learn the inference from the set of labelled instances used for training. The classification models are trained with the labelled instances; the new instances' class is predicted using that trained model when a new instance comes. Naïve Bayes algorithm and random forest are some of the examples of classification algorithms.

Another category of the machine learning algorithm is the unsupervised learning technique. In unsupervised learning, training takes place in the unlabelled data. Unsupervised learning uses only the input variable ($x_1$), and it has no corresponding output variable [9]. Based on the similarity of patterns, the outcome is predicted. Clustering is a kind of unsupervised learning technique. K-means algorithm and DBSCAN algorithm are some of the examples of unsupervised learning algorithms.

### 3.1 Classification Techniques

Classification is a supervised learning algorithm. Naïve Bayes algorithm and Hoeffding tree algorithm are some of the supervised learning algorithms that are used for the comparison in streaming and non-streaming environment. The classification model tries to learn the inference from the set of labelled instances used for training. The pre-trained model is used to predict the class label of the new instance.

### 3.2 Naive Bayes Classifier

Naïve Bayes algorithm is a kind of supervised machine learning classifier. It is the simplest and the popularly used classifier algorithm that is based on Bayes theorem. Naïve Bayes classifier assumes the occurrence of a particular feature in a class that is completely independent of the presence of any other feature. Naïve Bayes classifier is very simple and easy to build, and it is mainly very useful for larger datasets. It is called Naïve because it simplifies the assumptions on the attributes. The Naïve Bayes algorithm is called a probabilistic algorithm, and it is given by Eq. (2).

$$p\left(\frac{c_j}{x_j}\right) = \frac{p(x_j/c_j)p(c_j)}{p(x_j)} \tag{2}$$

$p\left(\frac{c_j}{x_j}\right)$ = Posterior probability gives the probability of $c_j$ occurring when given the evidence that $x_j$ has already occurred.

$p(c_j)$ = Prior probability of the class means that probability that $c_j$ occurs.

$p\left(\frac{x_j}{c_j}\right)$ = Prediction probability of a class means that probability of $x_j$ occurring given the evidence that $c_j$ has already occurred.

$p(x_j)$ = prior prediction probability of a class means that probability that $x_j$ occurs.

The error rate in the Bayes classifier is comparatively low when compared to other sophisticated models.

### 3.3 Hoeffding Tree Classifier

Hoeffding tree classifier is a kind of decision tree classifier. The decision tree algorithm is the most widely used algorithm. A decision tree is a tree structure that contains a root node, branches and leaf node that is used to predict the unlabelled instances. Hoeffding tree classifier is an incremental decision tree classifier that is capable of handling the massive data streams. In the traditional decision tree, the entire dataset is scanned to find the splitting criteria.

In the Hoeffding tree algorithm, instead of looking into the previously-stored instances, it waits and receives enough instances and then decides the splitting criteria for the root node [10]. Hoeffding tree entropy is given by Eq. (3)

$$\varepsilon = \sqrt{\frac{R^2 \ln / \delta}{2n_i}} \tag{3}$$

The parameter $R$ is the range of the random variable, $\delta$ is the desired probability of the estimate not being within ε of its expected value and n is the number of instances collected at the nodes [15].

## 3.4 Ensemble Classifier

Ensemble learning is used to improve the efficiency of the machine learning techniques by combining several base learners. This model gives a better efficiency when compared to the single classifier models. The ensemble technique is a kind of meta-algorithms that combines several ML techniques into a single predictive model that is used to decrease the variance, bias, and to improve the prediction accuracy of data [13]. The ensemble classifiers come under any one of the two categories, such as bagging or boosting. Resampling techniques are used to get different training samples for each classifier [14]. Bagging uses the concept of bootstrap samples, so it is named bootstrap aggregating. Bagging is used to generate multiple classifiers that are independent of one another [14]. In contrast, boosting assigns weights to the samples generated by the training dataset. Figure 2 shows the generalized ensemble model.

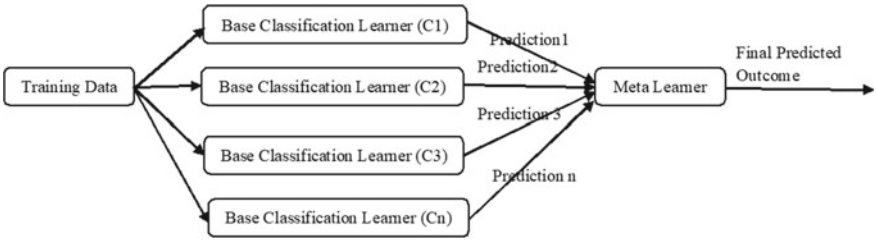



**Fig. 2** General ensemble classifier model. *Source* Wikimedia Commons

## 4 Attack Type Classification

Network and host-based attacks are becoming more prevalent nowadays. Hackers try to bypass security by misusing the already existing weaknesses in the network. The hackers disturb the network's normal functioning by sending a huge number of packets that floods the network, and they scan the network for vulnerabilities [4]. The attacker tries to find out the vulnerability in the network protocols or security devices like firewalls, intrusion detection systems and access to the network. The attackers can access the network or a host system by using sophisticated tools such as Nmap, Scapy, Metasploit, Net2pcap, etc. [4]

### *4.1 Data Description*

The dataset used for the analysis is the NSL_KDD dataset, the extended version of the KDD CUP 99 dataset. In the NSL_KDD dataset, all the duplicates are removed. Figure 3 shows the taxonomy of attack classification. This NSL_KDD dataset contains 41 features, of which 10 features are basic, while the next 12 features are called content features, and the rest 19 features are called traffic features [11].

Table 1 gives name of the attack and corresponding type of attack that are present in the NSL_KDD dataset based on the categories. Even though there are four different categories of attacks, namely denial of service (DOS) attack, probe attack, user to root (U2R) attack and remote to local (R2L) attack [12]. The NSL_KDD dataset has only two classes, normal or anomaly. The anomaly class encompasses the instances from all the four categories of attack.

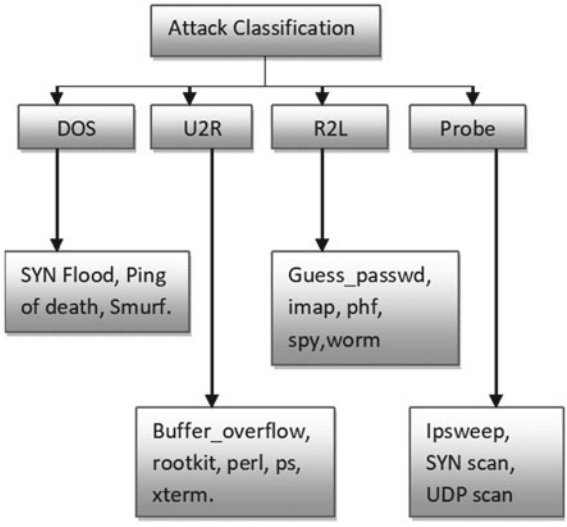



**Fig. 3** Taxonomy of attacks. *Source* Preeti Mishra [4]

**Table 1** List of attack based on attack categories

| Name of the attack | Type of attack |
|---|---|
| back | DOS attack |
| buffer_overflow | U2R attack |
| ftp_write | R2L attack |
| guess_passwd | R2L attack |
| imap | R2L attack |
| ipsweep | Probe attack |
| land | DOS attack |
| loadmodule | U2R attack |
| multihop | R2L attack |
| neptune | DOS attack |
| nmap | Probe attack |
| perl | U2R attack |
| phf | R2L attack |
| pod | DOS attack |
| portsweep | Probe attack |
| rootkit | U2R attack |
| satan | Probe attack |
| smurf | DOS attack |
| spy | R2L attack |
| teardrop | DOS attack |
| warezclient | R2L attack |
| warezmaster | R2L attack |

## *4.2 Denial of Service (DOS) Attack*

DOS attack makes the resources unavailable to the legitimate users by denouncing access to the information or services provided [5]. For example, a hacker can send several requests to access a specific resource or any running service instance. The server is flooded with several service requests and fails to offer genuine users the needed services. Smurf attack, ping to death attack, mail-bomb and SYN flood attack are examples of DOS attacks [4].

## *4.3 Probe Attack*

Probe attack on scanning gets to know the information about the network or a system. A probe attack is also called a scanning attack. Probe attack is one of the major issues in cyber security. The attacker uses the known vulnerabilities and sends a

large volume of scan packets to gain complete information about a network or a host. The attacker uses specialized tools such as Nmap, satan, saint, etc., to gain the information [4]. Depending on the way of gathering information, scanning can be active scanning or passive scanning. Ipsweep, SYN scan, UDP scan and reset scan are some of the examples of probe attacks.

### *4.4 User to Root (U2R) Attack*

User to root attack in which the exploits are used to gain the root access to the system by the unprivileged local user. Buffer overflow attack, Ffbconfig attack, Perl attack and Rootkits are some of the examples of the user to root attack. Rootkits give a hidden way for the hackers to escape the root privileges provided by the system. Rootkits allow hackers to hide suspicious processes by installing software, namely sniffer, keylogger, etc. [4]

### *4.5 Remote to Local (R2L) Attack*

Remote to local attack uses the group of exploits to get unauthorized access to the local machines. There are many ways in which the attacker gains legitimate privilege to the system. Dictionary/Guess password attack, FTP write attack, Xlock attack and Imap attack are some of the examples of remote to local attack [4].

## 5 Performance Evaluation

The algorithm's efficiency is measured using the performance measures like accuracy, precision, recall, and false alarm rate (FAR). An instance in an intrusion detection system can belong to either normal or anomaly. The confusion matrix is an N by N matrix used to measure the efficiency of machine learning classification models. The N specifies the number of instances in target classes. The actual targeted values are compared with the values predicted by the machine learning models. Figure 4 displays the confusion matrix used for the performance evaluation of the classification techniques.

The confusion matrix is given by four different values like true positive (TP), false positive (FP), false negative (FN) and true negative (TN).

**True Positive (TP)**: It is the number of actual attack instances that are appropriately classified as attacks.

**False Positive. (FP)**: It is the number of normal instances that are wrongly classified as an attack.

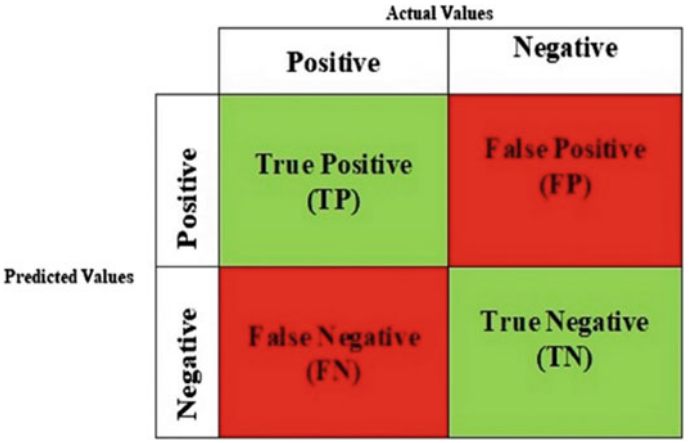



**Fig. 4** Performance evaluation using confusion matrix. *Source* Analytics Vidhya

**True Negative (TN)**: It is the number of normal instances that are appropriately classified as normal.

**False Negative (FN)**: It is the number of attack instances that are wrongly classified as normal.

**Accuracy (A)**

Accuracy is one of the performance measures used in machine learning techniques. Accuracy tells the number of instances correctly classified as normal or anomaly by an algorithm. Accuracy is given by Eq. (4).

$$A = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{4}$$

**Precision (P)**

Precision is one of the performance measures used in machine learning techniques. Precision tells how many of the correctly predicted instances actually turns out to be positive. Precision is given by Eq. (5)

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

**Recall (R)**

A recall is also one of the performance measures used in machine learning techniques. Recall tells how many of the actual positive cases can be predicted correctly by the model. The recall is given by Eq. (6)

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

**False Alarm Rate (FAR)**

False alarm rate is also one of the performance measures used in machine learning techniques. FAR tells how many correct instances are classified incorrectly as malicious instances. FAR is given by Eq. (7)

$$\text{FAR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{7}$$

## 6 Analysis of Machine Learning (ML) Techniques Performance on Intrusion Detection

The experiment is performed using the NSL_KDD benchmark dataset. The experiment is performed using the massive online analysis (MOA) framework, an open-source framework written in JAVA. MOA is mainly used to handle large data streams. MOA is mainly used to perform the analysis of static and dynamic data streams. The machine learning algorithms like Naïve Bayes and Hoeffding tree algorithm are used for the analysis. The performance of these algorithms is compared in both streaming and non-streaming environment. The performance metrics like accuracy, time and kappa values are compared, and inference is drawn.

The NSL_KDD dataset contains four different categories of attacks. All these categories are classified under a class called an anomaly, and another class is called normal. Figure 5 shows the number of instances under the normal and anomaly class. From the figure, it is clear that there is a balance between the normal and anomaly instances. This shows that the dataset used is not biased.

The heat map is the two-dimensional matrix representation of information. The colours are used to represent that information. A heat map is mainly used to represent both simple and complex information. The heat map helps visualize values concentration in finding the patterns and getting a deep perspective about the instances. The heat map is generated for all the feature variables. This is a powerful way to visualize the relationship between the variables in the high-dimensional space. Figure 6 shows the correlation heat map of the NSL_KDD dataset. This shows the relationship between the feature variables that are present in the NSL_KDD dataset. Darker colour takes higher values, and lighter colour takes lower values. This shows the correlation among the features in the dataset. The values in the heatmap range from -1 to 1. The value is close to 1 when the features are highly correlated, and value close to -1 shows the features are less correlated. The negative values in the heatmap show the features are negatively correlated, and the positive values are positively correlated. As the colour intensity increases that shows how strongly the variables

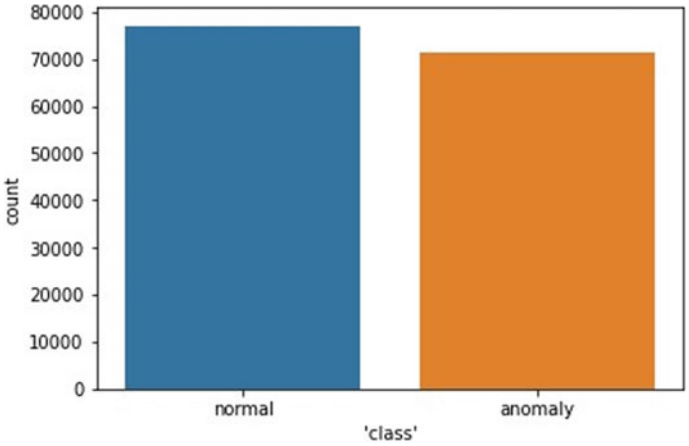



**Fig. 5** Counts of normal and anomaly instances of a dataset

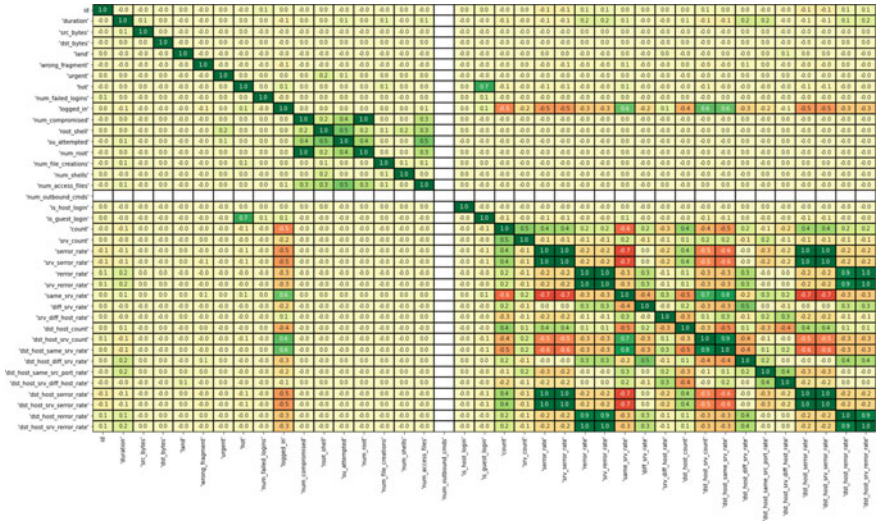


**Fig. 6** Heat map of NSL_KDD dataset

are correlated either it can be positively or negatively corelated. In Fig. 6, the feature 'num_outbound_cmd' shows empty white boxes with no values this means that removing this feature has no impact on others. The diagonal values of the matrix carry 1 except 'num_outbound_cmds and there are correlations between the same columns. Thus, from the heatmap, it is more evident that all the features are important in constructing the model.

The experimental setup uses the NSL_KDD dataset in a non-streaming and streaming environment. The static stream is generated using the MOA framework.

MOA is an open-source GUI-based framework designed in JAVA [15]. MOA handles stream data from static and evolving data streams. The comparison is made with three different categories of algorithms. The Naïve Bayes classifier is a probabilistic classifier, Hoeffding tree is a kind of decision tree classifier, and the Ensemble classifier. The whole dataset is divided into an 80:20 ratio. 80% of the data is used for training, and 20% of the data is used for testing. The algorithms are applied for the static dataset, static stream with and without concept drift. The comparisons are made based on performance generated by these algorithms. Figure 7 shows the sample screenshot of MOA environment that runs task concurrently and compares the results. MOA displays the results at the bottom of the GUI. It compares two different tasks. The red colour represents the current task, and blue colour represents the previous task.

The first scenario in which the accuracy of the dataset is compared between the three algorithms. It is found that the Hoeffding tree gives consistent accuracy in all three environments. From the graph in Fig. 8, it is clear that even in the streaming environment with concept drift, Hoeffding tree algorithm adapts to the change in the data and gives good accuracy. On the other hand, Naive Bayes performs well in the

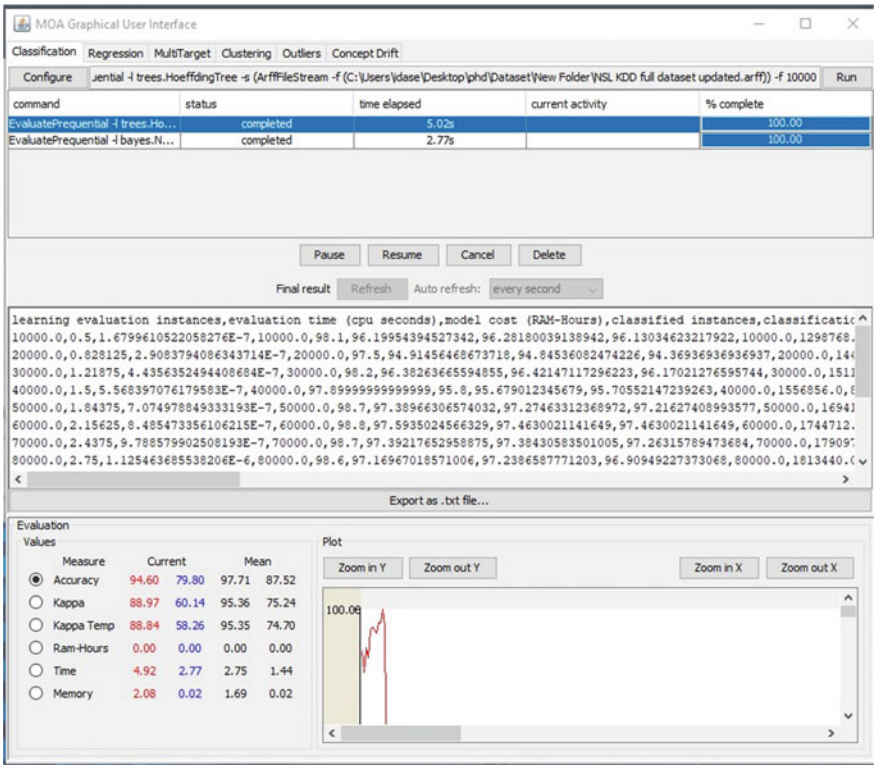



**Fig. 7** Snapshot of sample MOA framework

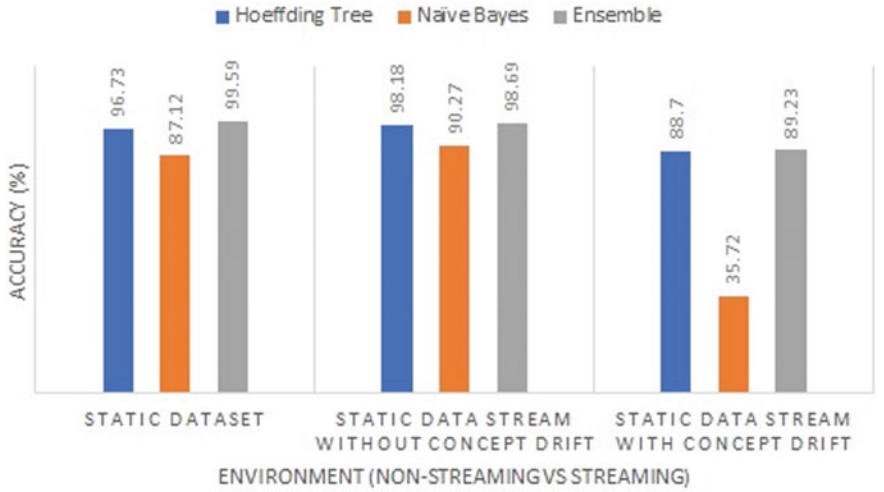



**Fig. 8** Accuracy in streaming and non-streaming environment

first two environments, but when the concept drift is induced in the static stream, it cannot withstand and adapt to the change. The detection accuracy of the Naïve Bayes classifier in the drifting stream gets drastically decreased. Compared to the Hoeffding tree and Naïve Bayes, ensemble classifiers give the highest accuracy in both streaming and non-streaming environments. This shows that ensemble classifiers are more suitable for stream data with and without concept drift. Figure 8 and Table 2 show the accuracy percentage obtained in both streaming and non-streaming environment. Table 2 shows that in a streaming environment with concept drift, the accuracy of Naïve Bayes is drastically reduced to 35.72%. This shows Naïve Bayes classifier is not able to adapt to the sudden changes that take place in the data. Hoeffding tree classifier and the ensemble classifier can adapt to the change and produce the detection accuracy of 88.7% and 89.23%. Even though the Hoeffding tree and ensemble classifier adapt to the drift, it gives more accuracy in the streaming environment without drift.

The second scenario in which the dataset time is taken to evaluate the instances is compared using three different algorithms: Hoeffding tree, Naïve Bayes and ensemble classifier. Figure 9 shows that the ensemble classifier takes more time

**Table 2** Accuraccy percentage in three different environment

| Accuracy (percentage) | Static dataset | Static data stream without concept drift | Static data stream with concept drift |
|---|---|---|---|
| Hoeffding tree | 96.73 | 98.18 | 88.7 |
| Naive bayes | 87.12 | 90.27 | 35.72 |
| Ensemble | 99.59 | 98.69 | 89.23 |

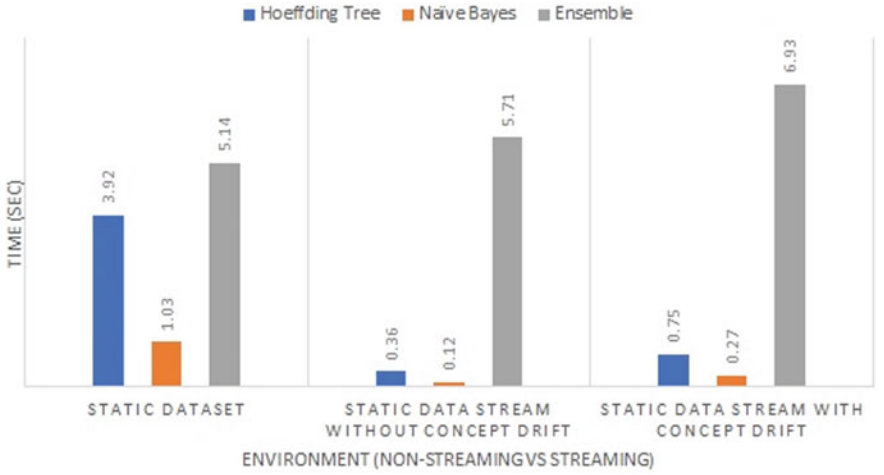



**Fig. 9** Time taken in non-streaming and streaming environment

than the Naïve Bayes algorithm and the Hoeffding tree algorithm. Ensemble classifiers use multiple algorithms and generate predictions that are combined to make the final prediction. It replaces the worst performance classifier and replaces it with a new classifier to improve the performance. Secondly, the Hoeffding tree takes more time because it is a kind of incremental algorithm. It waits for enough instances to perform the classification. Figure 9 and Table 3 show the time taken in the non-streaming and streaming environment. Table 3 shows that the Naïve Bayes classifier takes less time to give its outcome in all three environments. The ensemble classifier takes a maximum of 6.93 s in a streaming environment with concept drift. Naïve Bayes and Hoeffding tree classifier take 0.27 and 0.75 s. It is evident that ensemble classifier takes more time for computation but gives more detection accuracy.

In the third scenario, the dataset is taken, and the kappa value given by the three different algorithms in both the environments are compared. From Fig. 10, it is found that the ensemble algorithm gives more kappa value when compared to other algorithms. It shows that the agreement level between the raters is more in the ensemble classifier algorithm. When the accuracy is more automatically, the kappa value produced increases. In the Naïve Bayes classifier, the kappa value is zero. When the Kappa value is zero, it is obvious that there is no agreement among the

**Table 3** Time taken in three different environments

| Time (S) | Static dataset | Static data stream without concept drift | Static data stream with concept drift |
|---|---|---|---|
| Hoeffding tree | 3.92 | 0.36 | 0.75 |
| Naive bayes | 1.03 | 0.12 | 0.27 |
| Ensemble | 5.14 | 5.71 | 6.93 |

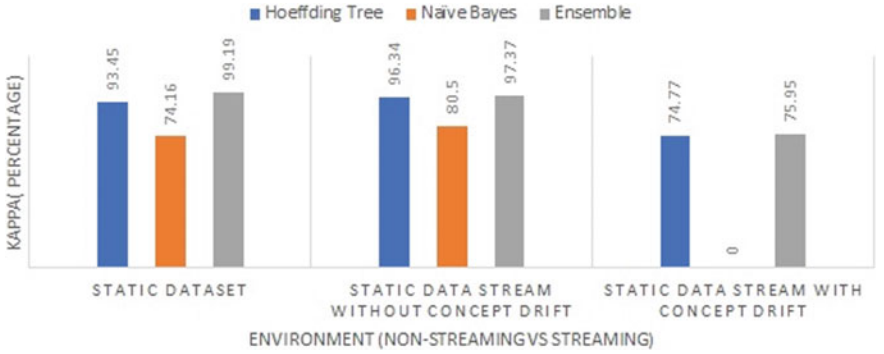



**Fig. 10** Kappa value in non-streaming and streaming environment

**Table 4** Kappa values in three different environment

| Kappa (percentage) | Static dataset | Static data stream without concept drift | Static data stream with concept drift |
|---|---|---|---|
| Hoeffding tree | 93.45 | 96.34 | 74.77 |
| Naive bayes | 74.16 | 80.5 | 0 |
| Ensemble | 99.19 | 97.37 | 75.95 |

raters. Naïve Bayes' accuracy gets drastically reduced; thereby, it reduces the kappa values. If the kappa value is less than or equal to zero, there is no adequate agreement among the raters. Figure 10 and Table 4 show the kappa value given by the three algorithms in all three environments. Table 4 shows that Naïve Bayes classifier kappa values are less in all three environments than the Hoeffding tree and ensemble classifier. It is very clear from the table that Naïve Bayes produces zero in a streaming environment with concept drift. This shows that as the accuracy reduces drastically, the kappa value also reduces. Zero kappa values show that there is no agreement between the raters. Kappa values produced by the Hoeffding tree and ensemble classifier in the drift stream environment are 74.77 and 75.95%. Kappa value is more in non-streaming and streaming environments without concept drift. If the kappa value is less than 60%, it indicates an inadequate agreement among the raters.

## 7 Conclusion

This paper discusses the performance of three different algorithms that are used in three different environments like non-streaming, streaming with concept drift and streaming without concept drift. The Naïve Bayes classifier is a kind of probabilistic classifier. The Hoeffding tree classifier is the kind of decision tree incremental classifier and ensemble classifiers. The three different classifiers used for the comparison

found that ensemble classifiers give consistent performance in all three environments than other algorithms. The time consumed to give the outcome is more in ensemble classifiers. The competitive study shows that the Naïve Bayes classifier, the probabilistic classifier, is not suitable for the streaming environment with concept drift. This means that the Naïve Bayes classifier cannot adapt to the changes induced in the concepts. On the other hand, Hoeffding tree and ensemble classifiers can adapt to the concept drift induced in the stream. The concept drift is induced in the static stream by using SEA generator. Thus, from the competitive study is performed, ensemble classifiers are found to be more suitable in non-streaming and streaming environments. Though ensemble classifier is more suitable in all the tree environment, still the detection accuracy in drift stream is less. In future, the detection accuracy in drift stream environment can be improved.

## References

1. Ambusaidi MA, He X, Nanda P, Tan Z (2016) Building an intrusion detection system using a filter-based feature selection algorithm. IEEE Trans Comput
2. Shone N, Ngoc TN, Phai VD, Shi Q (2018) A deep learning approach to network intrusion detection, IEEE Trans Emerging Topics Comput Intell 2
3. Green C, Lee B, Amaresh S, Engels DW (2018) Comparative study of deep learning models for network intrusion detection. SMU Data Sci Rev 1
4. Mishra P, Varadharajan V, Tupakula U, Pilli ES (2018) A detailed investigation and analysis of using machine learning technique for intrusion detection, IEEE Commun Surv Tutorials
5. Mantas G, Stakhanova N, Gonzalez H, Jazi HH, Ghorbani AA (2015) Application-Layer denial of service attacks: taxonomy and survey. Int J Inf Comput Secur
6. Peng K, Leung VCM, Huang Q (2018) Clustering approach based on mini batch kmeans for intrusion detection system over big data. Spec Section Cyber Phys Soc Comput Netw
7. Pham NT, Foo E, Suriadi S, Jeffrey H, Lahza HFM (2018) Improving performance of intrusion detection system using ensemble methods and feature selection, ACMW, Brisbane, Australia
8. Desale KS, Kumathekar CN, Chavan AP (2015) Efficient intrusion detection system using stream data mining classification technique. In: International conference on computing communication control and automation. IEEE Comput Soc
9. Haq NF, Rafni M, Onik AR, Shah FM, Hridoy MAK, Farid DM (2015) Application of machine learning approaches in intrusion detection system: a survey. Int J Adv Res Artif Intell
10. Muallem A, Shetty S, Pan JW, Zhao J, Biswal B (2017) Hoeffding tree algorithms for anomaly detection in streaming datasets: a survey. J Inf Secur
11. Yin C, Zhu Y, Fei J, He X (2017) A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access
12. Seraphim BI, Palit S, Srivastava K, Poovammal E (2019) Implementation of machine learning techniques applied to the network intrusion detection system. Int J Eng Technol
13. Smolyakov V (2017) Ensemble Learning to improve machine learning results, cube.js open source analytics framework
14. Gopika D, Azhagusundari B (2014) An analysis on ensemble methods in classification tasks. Int J Adv Res Comput Commun Eng
15. Bifet A, Gavalda R, Holmes G, Pfahringer B (2017) Machine learning for data streams with practical examples in MOA. The MIT Press
16. Srimani PK, Patil MM (2016) Mining data streams with concept drift in massive online analysis frame work. Wseas Trans Comput 15

17. Hamid Y, Balasaraswathi VR, Journaux L, Sugumaran M (2018) Benchmark datasets for network intrusion detection: a review. Int J Netw Secur
18. Wang L (2017) Big data in intrusion detection systems and intrusion prevention system. J Comput Netw

# Content Related Feature Analysis for Fake Online Consumer Review Detection


**Dushyanthi Udeshika Vidanagama, Thushari Silva, and Asoka Karunananda**

**Abstract** Due to the advancements of Internet, majority of people will prefer to buy and sell the commodities through e-commerce websites. In general, people mostly trust on reviews before taking the decisions. Fraudulent reviewers will consider this as an opportunity to write fake reviews for misleading both the customers and producers. There is a necessity to identify fake reviews before making it available for decision making. This research focuses on fake review detection by using content-related features, which includes linguistic features, POS features, and sentiment analysis features. Ontology-based method is used for performing the aspect-wise sentiment analysis. All the features of reviews are calculated and incorporated into the ontology, and fake review detection is also accelerated through the rule-based classifier by inferencing and querying the ontology. Due to the issues related with labeled dataset, the outliers from an unlabeled dataset were selected as fake reviews. The performance measure of the rule-based classifier outperforms by incorporating all the content-related features.

**Keywords** Ontology · Rule-based classifier · Outliers · Content-related features

D. U. Vidanagama (✉) · T. Silva · A. Karunananda
Department of Computational Mathematics, Faculty of Information Technology, University of Moratuwa, Moratuwa, Sri Lanka
e-mail: udeshika@kdu.ac.lk

T. Silva
e-mail: thusharip@uom.lk



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_35

## 1 Introduction

With the vast advancements of Internet, majority of people are preferring online shopping. As there is no limitation or user restriction over the Internet and E-commerce, anyone can do buying and selling through e-commerce websites and write opinions against the products/services itself. These opinions highlight positive or negative aspects of online product/services which can be used for further purchasing decisions and product improvement decisions. But most of the time people do not read the whole reviews, but just look at the individual rating, overall rating or aspect-wise summaries of the particular product/service. Hence, people are highly focused on automatic analysis of reviews by deploying sentiment analysis rather than manual reading.

Sentiment analysis is very essential for producers and service providers as it monitors the reviews for numerous aspects, which led them to enhance product quality and service improvements. Some fraudulent reviewers consider this as an opportunity to write fake/spam reviews for misleading customers and producers as those will affect the sentiment produced by the reviews. Fraudulent reviewers may write fake positive reviews to promote or fake negative reviews for demote product's reputation. Customers may get negative impression on product/services if they were deceived by fake reviews. Henceforth, it is very essential to detect fake reviews before applying sentiment analysis techniques. This research uses the terms 'opinion spam,' 'fake reviews,' and 'spam reviews' interchangeably. The fake review detection can be considered as a significant process of sentiment analysis. It is a challenging task which has a close relationship with sentiment analysis.

The reviewers can express their opinions within the body of the review as well as they can express the overall sentiment about the product/service as a rating. But most of the time, the rating does not clearly represent the sentiment orientation of the review. It is argued that user rating should consistent with the tone of the review text published by the same user [1]. According to the observations regarding the rating score, the content of the reviews indicated the existence of large number of reviews whose rating and the sentiment score are inconsistent [2, 3]. Therefore, it can be argued that the content of the review is more important than the rating of the review. The major contribution of this paper is to incorporate sentiment orientation and sentiment of the reviews along with some other review content-related features for fake review detection. Although there are different methods for deceptive review detection, this research mainly focuses on sentiment orientation and sentiment of the review. So, it is important to find the sentiment orientation and total sentiment of the reviews using feature-based sentiment analysis methods.

The rest of the paper is organized as follows. Section 2 discusses the related research work in the area of fake review detection. Section 3 includes the proposed approach of feature-based sentiment analysis method for fake review detection. Section 4 includes the experiment and evaluation of the fake review detection method on the Amazon unlocked mobile review dataset. Finally, Sect. 5 concludes the paper with conclusion and future research directions.

## 2 Related Work

The sentiment analysis is to represent the attitude of the reviewer with respect to a certain aspect or overall polarity of the review. The basic task of sentiment analysis is to categorize the polarity of the review at sentence level, document level, or aspect/feature level into positive, negative, or neutral [4]. The aspect level sentiment analysis considers the related aspects of each sentence and determines the overall polarity of the review. Features or its synonyms are explicit features, while the implicit features are not directly appear in the review, but may contain some aspect about the product. Ngoc [5] used a conditional probabilistic model combined with the bootstrap technique to extract important aspect words. The core aspects are then enlarged by inserting words that have high probability to appear in the same sentences that they occur. Fang and Zhan [6] proposed the sentiment polarity categorization by negation phrase identification and sentiment score computation. The sentiment polarity categorization was performed at sentence level and review level. The sentiment score of sentiment phrases was calculated based on the number of occurrences of each sentiment token in each star rating separately. Zhang et al. [1] also used aspect-based sentiment analysis method where the aspects of the products are grouped into categories. The semantic-based method was used to identify explicit aspects, and point-wise mutual information (PMI)-based statistical method was used to identify the implicit features by calculating the co-occurrence frequency of potential feature and actual feature. Apart from these methods, some research work used ontology-based methods for feature level sentiment analysis [7–9].

Since all the reviews are not genuine, it is essential to detect fake reviews before having the sentiment analysis and summarization. The review spams are categorized into three types including untruthful reviews (type 1), brand only reviews (type 2), and non-reviews (type 3) [10]. Generally, the non-reviews contain unwanted content such as advertisements, links, email addresses, phone numbers, prices, etc. The high percentage of entities from brand only reviews belong to product name, manufacturer name, distributer, country names, etc. The type 2 and type 3 kind of spam reviews can be easily detected and remove before the summarization. The type 1 kind of untruthful reviews need to be analyzed carefully for identification. This research focuses on detecting all the three types of fake reviews before making the sentiment summary.

The review content-related features, reviewer behavior applied using machine learning techniques, and lingustic patterns are used by using supervised learning algorithms [11]. Rout et al. [12] demonstrated how the popular semi-supervised learning approaches such as expectation maximization algorithm, co-training algorithm, label propagation, and this used for the linguistic applications. For unsupervised learning methods [13, 14] used Bayesian clustering method in which their model adopted the spamming degress of authors and reviews using latent spam model (LSM). Further, [15] used the unsupervised $K^{th}$ nearest neibour (KNN) method, and this method mainly focused on the detection review outliers and its features. Apart from these machine learning approaches, Wang et al. [16] used review graph-based approach which helps to capture the fake reviews.

Further [17] introduced an approach that maps the singleton review spam detection problem to an abnormally correlated temporal pattern-detection problem. Li et al. [18] also considered meta-data patterns, both temporal and spatial when detecting untruthful reviews. Apart from these methods, some authors used outlier detection methods to categorize the reviews as spam or truthful [12, 19].

Feature extraction is always considered as a critical step in the process of identifying opinion spam. The effectiveness and accuracy of any algorithm depend on the input features provided to that algorithm [20]. Many authors investigated for linguistic features, POS styles, writing patterns when detecting deceptive reviews [21–23]. Apart from linguistic features, most of the researchers proposed spam detection by incorporating rating deviation, content similarity, and inconsistency among rating and the sentiment [3, 24, 25].

Most of the supervised and semi-supervised approaches used for fake review detection may have the following limitations [26]. Firstly, the manual labeling dataset with fake and non-fake reviwes always cannot be considered as authentic. Secondly, the scarcity of labeled spam review data for model learning. Also it was proved that fake reviews which are generated through crowd sourcing mechanisms or pseudo fake reviews may not be valid training data because the models do not generalize well on real-time test data [14, 20]. Due to limitation of labeled dataset creation, it is necessary to find a mechanism to create a gold standard dataset with fake reviews without using manual labeling dataset for model training.

## 3 Proposed Methodology

The proposed methodology for detecting fake reviews is shown in Fig. 1. It consists with three main modules: (1) Pre-Processing (2) Extraction, (3) Ontology Management, and (4) Spam Detection. This approach can be used by any E-commerce-related website to detect fake reviews at the time of posting a review. When a new review is posted, it can be gone through these modules and finally resulted whether the particular review is fake/truthful. The following subsections explained each module in advance.

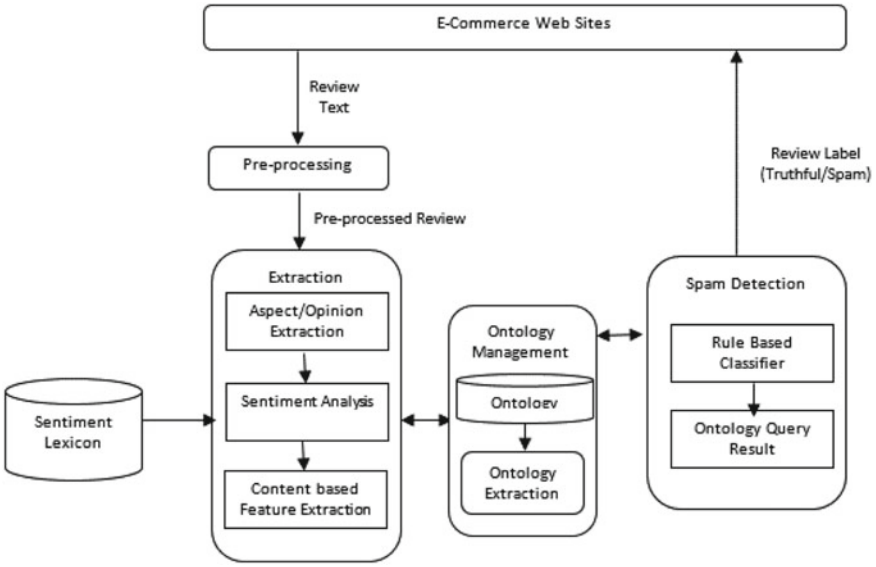



**Fig. 1** Proposed fake detection model

Pre-processing module is responsible for removing unwanted words and characters, sentence splitting, spell correction, stop words removal, tokenization, POS tagging, lemmatization, and dependency parsing. The pre-processed text which is the output of this module is important for the extraction module. The Python libraries are used for the tasks of this module.

The aspect extraction process is accelerated by mapping the extracted aspects with the aspect-related concepts of a domain ontology. The purpose of the ontology is to identify the fake reviews in a specific domain. The sentiment analysis and filtering of fake reviews are done through the ontology in the specific domain. By replacing the ontology, the fake reviews can be filtered in any domain. The ontology is developed in Protégé software. This ontology consists with some concepts like feature, OpinionWords, review, product, and sentence. The feature class consists with subclasses of specific features of the product itself. It has a data property on sentiment expectation of each aspect which contains an integer of either −1 or 1. OpinionWords class contains the instances of opinion words. Review class contains the instances of reviews. Sentence class contains the instances of sentences of the reviews. All the classes have object properties and data properties for different purposes. The ontology can be updated with newly identified features for accurately perform the sentiment analysis. The instances and related properties are updated according to the arrival of new reviews.

The extraction module comprises with two processes: Aspect and opinion extraction and content—based feature extraction.

### 3.1 Aspects and Opinion Extraction

Every time people express their opinion on a product/service aligned with the aspects of the product/service. The associated aspects and the opinions are extracted within this process. Always the aspects may be a noun or a noun phrase [11]. By following a rule-based strategy which are derived using the dependency parsing and POS tag patterns, the candidate aspects and candidate opinions can be extracted. All the extracted candidate aspects are nouns/noun phrases, but most of the extracted candidate nouns/noun phrases may not related to the domain. Therefore, to reduce the number of candidate aspects which are going to map with the ontology, an aspect score value is calculated (1). The aspect score value is calculated using an existing domain specific corpus of reviews. Here, $x$ is the candidate aspect, $y_i$ is the existing aspects of the ontology, $f(x, y_i)$ is the number of times that the feature $x$ and $y_i$ are co-occurred in each sentence, $f(x)$ is the number of sentences in the corpus where $x$ appears, and $N$ is the total number of sentences in the corpus.

$$\text{Aspect} - \text{Score} = \sum_i \frac{f(x, y_i)}{f(x) f(y_i)} * N \tag{1}$$

For the candidate aspects who have the aspect score value greater than one can be selected as appropriate features, if not the candidate aspect is disregarded. Firstly, the selected candidate aspects are mapped with the aspect concepts of the ontology. Secondly, if such an aspect cannot be mapped with the ontology concept, all the possible synonyms are to be extracted from WordNet database [27] and map the synonym with the ontology concepts. Thirdly, if any selected candidate aspect which cannot be mapped to ontology concept from first and second method, then that aspect needs to be updated under a selected subcategory which has gain high PMI value among all the subcategories (2). Here, x is the candidate selected feature, $S_i$ is the $i$th subcategory among the features of the ontology, and N is the number of review sentences.

$$\text{PMI}(x, S_i) = \log_2 \frac{\text{GoogleHitsCount}(x, S_i)}{\text{GoogleHitsCount}(x) \times \text{GoogleHitsCount}(S_i)} \times N \tag{2}$$

The opinions are expressed against the aspects of the product/service. After extracting the aspects from the reviews, the opinion extraction process has to follow a rule-based strategy based on the POS tag pattern and the type of the dependency relation which is identified from Stanford Dependency Parser [28]. Mostly the opinion may be an adjective, verb, or adverb. Then the extracted candidate opinions are compared against the SentiWordNet [29] for existence and selected as relevant opinion words associated with the aspect (ex: very, quickly, extremely, slowly, etc.) or negation words (ex: no, never, nothing, without, etc.) or conjunction words (ex: but, and, or, etc.) which may change the sentiment of the opinion. These kinds of adverbs

```
Input: For each aspect-opinion pair - Aspect (A), Opinion (O), Adverb (adv), number of negation
words (n), Sentiment Score of opinion word - SC(O)
Output: Sentiment score of aspect – opinion pair – SS(A-O)
If adv does not exist, Then
 SS(A-O) = (-1)^n * SS (O) * SE (A)
Else
  If SS (adv) > 0, then
     SS(A-O) = (-1)^n * min {1, SS (adv)+SS(O)} * SE (A)

  If SS (adv)< 0, then
     SS(A-O) = (-1)^n * max { -1, SS (adv)+SS(O)} * SE (A)

End If
```

**Fig. 2** Algorithm for sentiment-score calculation for aspect-opinion pairs

are used to modify a verb, adjective, or another adverb. The next process of the extraction module is to calculate the sentiment value of each extracted aspect-opinion pair and aggregate them to generate the overall sentiment of each review.

The sentiment score of each aspect-opinion pair is calculated using the algorithm in Fig. 2. The sentiment score of the opinion word is retrieved from the sentiment lexicon database, SentiWordNet. As there are three score values of each synsets of SentiWordNet, the maximum out of the positive or negative score is retrieved as the sentiment score value of the opinion word. After calculating all the score values of pairs in each review, the overall positive sentiment score and overall negative sentiment score are calculated using (3) and (4).

$$\text{Sentiment}_{\text{Positive}}(\text{review}) = \min\left\{1, \sum_{i}^{n} \text{SS}_i(A - O)\right\} \tag{3}$$

$$\text{Sentiment}_{\text{Negative}}(\text{review}) = \max\left\{-1, \sum_{j}^{m} \text{SS}_j(A - O)\right\} \tag{4}$$

where $n$ is the number of positive aspect-opinion pairs and m is the number of negative aspect-opinion pairs. The overall sentiment of the review is generated using (5).

$$\begin{aligned}\text{Sentiment (review)} = {} & \text{Sentiment}_{\text{Positive}}(\text{review}) \\ & + \text{Sentiment}_{\text{Negative}}(\text{review})\end{aligned} \tag{5}$$

The general rating value of five-point polarity system is then normalized into [−1, 1] as the calculated sentiment polarity value is aligned with a scale where −1 represents extreme negative and +1 represents extreme positive.

### *3.2 Content-Based Feature Extraction*

This process generates a set of review content-related features which are used to identify the fake reviews and truthful reviews. Table 1 shows the descriptions about the extracted content-based features. The sentiment analysis related features should be extracted from the previous aspects and opinion extraction process.

### *3.3 Fake Detection Module*

For some of the features, threshold values are defined for making the final decisions. Those features are compared with the threshold values when deciding on the fake or truthfulness. Based on the comparisons over the content-based features, the review can be classified as truthful or spam. The content-related features of the incoming review are calculated and stored within the ontology. Once inferencing, the review is automatically classified into the spam or truthful class according to the feature values and rules. By querying, all the available spam reviews can be displayed (Fig. 3), while the class category of a particular review can also be retrieved (Fig. 4).

## 4 Experiment and Evaluation

Most of the supervised and semi-supervised fake detection methods used labeled dataset for the model training. It was difficult to find a huge amount of labeled dataset which can guarantee the accuracy. Therefore, this research used unlabeled dataset of Amazon unlocked mobile reviews which is provided by Kaggle for model training. This unlabeled dataset contains 5000 unlabeled reviews with the review content and the rating. The variation of the given rating by the reviewers and the calculated sentiment polarity of the dataset is shown in Fig. 5. Generally, the outliers are deviated from common reviews, and it can be argued that the outliers may be generated through some other mechanism which is not consistent with general reviews. This argument is used to generate labeled spam reviews as outliers. The Mahalanobis distance [30] is used to detect the outliers of the multivariate dataset. The outlier distribution according to the existing opinion words of reviews is shown in Fig. 6.

**Table 1** Content-related features

| Dimension | Content—based features | Description |
|---|---|---|
| Linguistic features | Total word count (F1) | Total number of words in the review |
| | Average word count (F2) | Average number of words per sentence |
| | Ratio of numeral word count (F3) | Total number of words containing numerical values/Total word count |
| | Ratio of exclusive words (F4) | Total number of exclusive words/Total word count |
| | Ratio of negation words (F5) | Total number of negation words/Total word count |
| | Ratio of causation words (F6) | Total number of causation words/Total word count |
| | Ratio of capital letters (F7) | Total number of words containing capital letters/Total word count |
| | Ratio of all caps word (F8) | Total number of all caps words/Total word count |
| | Content similarity (CS) (F9) | The similarity percentage of review text to other existing review contents |
| | Ratio of exclamation marks (F10) | Total number of words with exclamation marks/Total word count |
| | Ratio of question marks (F11) | Total number of words with question marks/Total word count |
| | Ratio of named entities (F12) | Total number of named entities/Total word count |
| POS features | Ratio of nouns (F13) /adjectives (F14)/prepositions (F15)/determiners (F16)/verbs (F17)/adverbs (F18)/connector words (F19)/first person pronouns (F20)/pronouns (F21) | Total number of POS words/Total word count |
| Dimension | Content—based features | Description |
| Sentiment analysis | Ratio of aspect-related words (F22) | Total of number of aspects/Total words |
| | Ratio of opinion words(F23) | Total number of opinions/Total words |
| | Ratio of positive (F24) /negative (F25) opinion terms | Number of positive/negative opinion terms/Total opinion count |
| | Sentiment deviation (SD) (F26) | Difference between the calculated sentiment score and user rating |

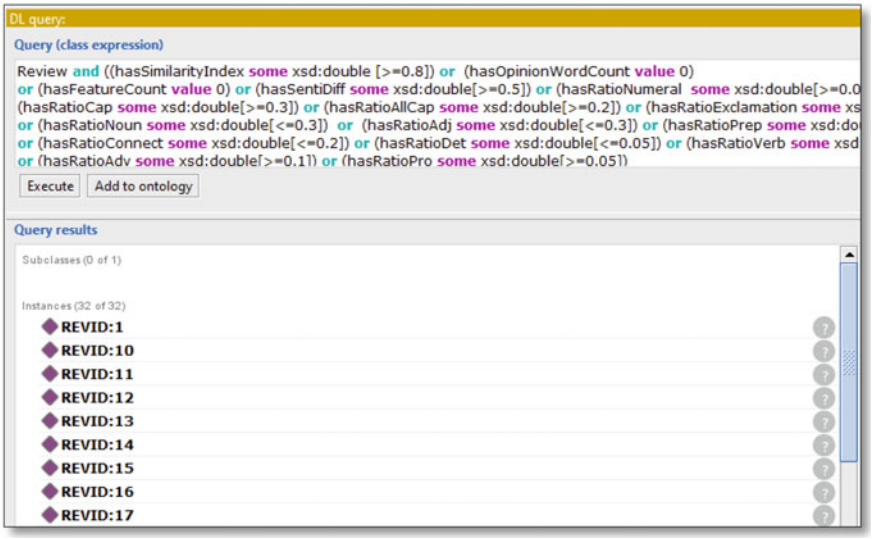



**Fig. 3** Results for available spam reviews in ontology

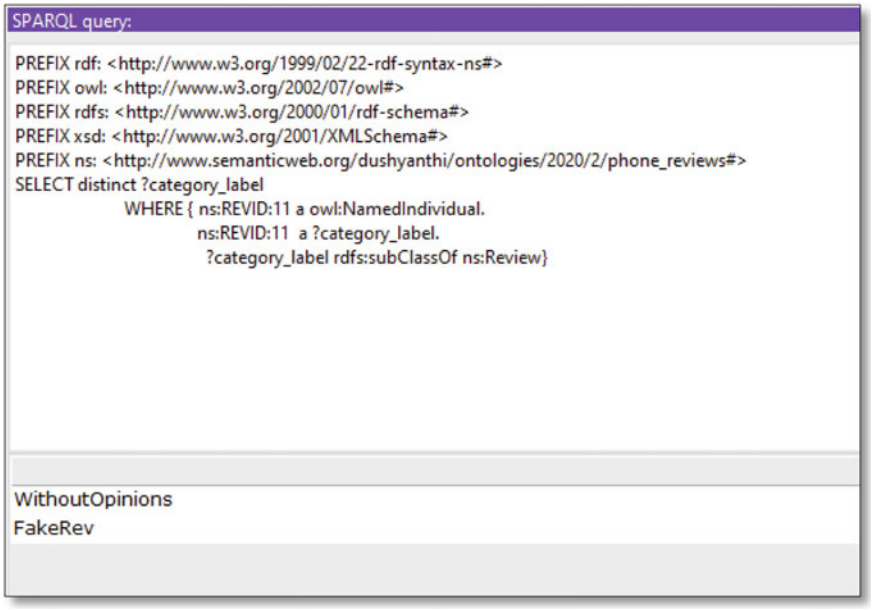



**Fig. 4** SPARQL result of the category of review

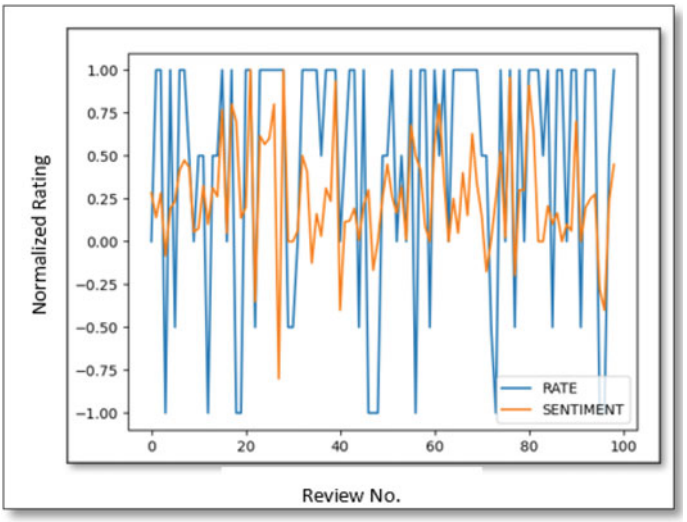



**Fig. 5** Variation of given rating and sentiment polarity

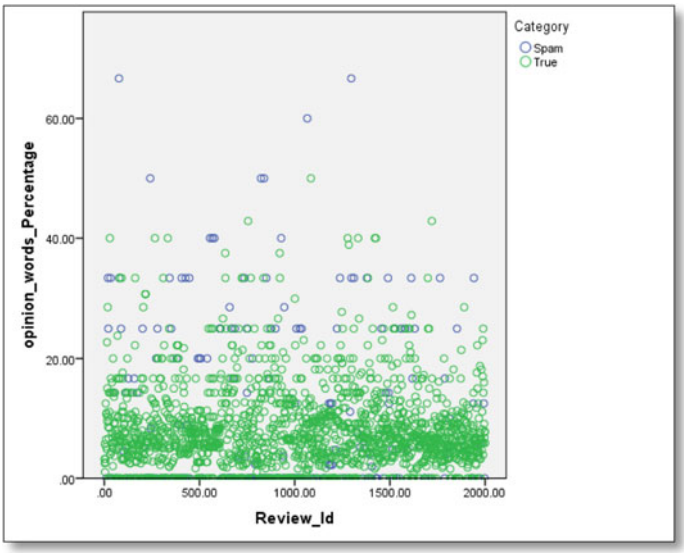



**Fig. 6** Outlier distribution over ratio of opinion words

The calculated content-related features are used to cluster the dataset for identifying the spam and truthful labeled data. Then, a labeled dataset was selected with 1500 spam reviews and 1500 truthful reviews, where 70% from the dataset is then

used as training sample, 15% is used for testing, and the other part is used for validation. This training dataset is then used for finding the threshold values of the classifier rules. An iterative approach is taken to fine-tune the threshold values. Finally, the modified rules are included in the rule engine of protégé as SWRL rules. When a new review is arrived, it will be stored in the ontology as a review instance and calculate the content-related feature values. These values are stored as data properties of the ontology. After inferencing, the SPARQL query result will provide the classification result on a review saying whether it is spam or truthful based on rules of the rule engine of the ontology.

In order to evaluate the effect of linguistic features, POS features, and sentiment analysis features, the rule-based classifier is used. As shown in Table 2, different combinations of review content-related features are used to evaluate the performance of the classification. When considering linguistic features only and by combination of linguistic features and POS features, the classifier reached the accuracy of 64.1% and 73.3%, respectively. But when considering linguistic features, POS features and sentiment-related features as a combination, the classifier reached to high values of precision, recall, accuracy, and F1 score.

## 5 Conclusion and Further Work

The proposed research work has extensively investigated the spam review classification method by incorporating the review content-related features without considering a labeled dataset. The outliers of an unlabeled dataset were considered as spam reviews, and the rest of the reviews were used as truthful reviews for training the model. A rule-based classifier was used for the classification. The content-related features such as linguistic features, POS features, and sentiment analysis features are considered when performing the classification. The sentiment analysis is performed by using the domain-related aspects of the ontology. All the features are calculated for each new review and stored within the ontology. When inferencing, the classification result will be generated according to the SWRL rules of the ontology. The accuracy, recall, precision, and F1 score values of the rule-based classifier are slightly increased by incorporating all the linguistic features, POS features, and sentiment analysis features together without considering them individually. Further, the research directions will be suggested to incorporate the reviewer-related features along with the review content-related features and explore the performance of different classifiers.

**Table 2** Performance metrics of classification results

| Feature Combination | Precision (%) | Recall (%) | Accuracy (%) | F1 Score (%) |
|---|---|---|---|---|
| Linguistic | 61.11 | 71.6 | 64.1 | 65.9 |
| Linguistic + POS | 68 | 70.2 | 73.3 | 69.1 |
| Linguistic + POS + Sentiment | 88.69 | 90.90 | 88.98 | 89.79 |

**Acknowledgements** The authors would like to thank all the staff members and family members who gave their maximum support during the completion of this research.

## References

1. Zhang K, Cheng Y, Xie Y, Honbo D, Agrawal A, Palsetia D, Lee K, Liao W, Chaudhari A (2001) SES: sentiment elicitation system for social media data. In: 2011 IEEE 11th international conference on data mining workshops, Vancouver, BC, Canada, pp 129–136. https://doi.org/10.1109/icdmw.2011.153
2. Ganeshbhai SY, Shah BK (2015) Feature based opinion mining: a survey. In: 2015 IEEE international advance computing conference (IACC), Banglore, India, pp 919–923. https://doi.org/10.1109/iadcc.2015.7154839
3. Peng Q, Zhong M (2014) Detecting spam review through sentiment analysis. J Softw 9(8):2065–2072. https://doi.org/10.4304/jsw.9.8.2065-2072
4. Pawar MS (2015) Formation of smart sentiment analysis technique for big data. Int J Innov Res Comput Commun Eng IJRCCE 02(12):7481–7488. https://doi.org/10.15680/ijircce.2014.0212034
5. Ngoc TNT, Nguyen T (2019) Mining aspects of customer's review on the social network. J Big Data 21
6. Fang X, Zhan J (2015) Sentiment analysis using product review data. J Big Data 2(1):5. https://doi.org/10.1186/s40537-015-0015-2
7. Agarwal B, Mittal N, Bansal P, Garg S (2015) Sentiment analysis using common-sense and context information. Comput Intell Neurosci 2015:1–9. https://doi.org/10.1155/2015/715730
8. Salas-Zárate MP, Valencia-García R, Ruiz-Martínez A, Colomo-Palacios R (2017) Feature-based opinion mining in financial news: an ontology-driven approach. J Inf Sci 43(4):458–479
9. Wójcik K, Tuchowski J (2014) Ontology based approach to sentiment analysis. In: 6th International scientific conference faculty management crackonomics university economy, p 14
10. Jindal N, Liu B (2008) Opinion spam and analysis. In: Proceedings of the international conference on Web search and web data mining—WSDM '08, Palo Alto, California, USA, p 219. https://doi.org/10.1145/1341531.1341560
11. Vidanagama DU, Silva TP, Karunananda AS (2020) Deceptive consumer review detection: a survey. Artif Intell Rev 53(2):1323–1352. https://doi.org/10.1007/s10462-019-09697-5
12. Rout JK, Dalmia A, Choo KKR, Bakshi S, Jena SK (2017) Revisiting semi-supervised learning for online deceptive review detection. vol 5, pp 1319–1327. IEEE Access. https://doi.org/10.1109/access.2017.2655032
13. Mukherjee A, Venkataraman V, Liu B, Glance N (2013) What yelp fake review filter might be doing? In: Proceedings of the ınternational AAAI conference on web and social media, p 10
14. Lin Y, Zhu T, Wang X, Zhang J, Zhou A (2014) Towards online review spam detection. In: Proceedings of the 23rd international conference on World Wide Web—WWW '14 Companion, Seoul, Korea, pp 341–342. https://doi.org/10.1145/2567948.2577293
15. Singh S (2015) Improved techniques for online review spam detection. MTech Natl Inst Technol 58
16. Wang G, Xie S, Liu B, Yu PS (2011) Review graph based online store review spammer detection. In: 2011 IEEE 11th international conference on data mining, Vancouver, BC, Canada, pp 1242–1247. https://doi.org/10.1109/icdm.2011.124
17. Xie, S, Wang, G, Lin S, Yu PS (2012) Review spam detection via temporal pattern discovery. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining—KDD '12, pp 823, Beijing, China (2012). https://doi.org/10.1145/2339530.2339662

18. Li H, Chen Z, Mukherjee A, Liu B, Shao J (2015) Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns. In: Proceedings of ninth international AAAI conference on web social media, p 4
19. Liu W, He J, Han S, Cai F, Yang Z, Zhu N (2019) A method for the detection of fake reviews based on temporal features of reviews and comments. IEEE Eng Manag Rev 47(4):67–79. https://doi.org/10.1109/EMR.2019.2928964
20. Rastogi A, Mehrotra M (2017) Opinion spam detection in online reviews. J Inf Knowl Manag 16(4):1750036. https://doi.org/10.1142/s0219649217500368
21. Newman ML, Pennebaker JW, Berry DS, Richards JM (2003) Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull 29(5):665–675. https://doi.org/10.1177/0146167203029005010
22. Li J, Ott M, Cardie C, Hovy E (2014) Towards a general rule for identifying deceptive opinion spam. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, vol 1, pp 1566–1576, Baltimore, Maryland. https://doi.org/10.3115/v1/p14-1147
23. Ong T, Mannino M, Gregg D (2014) Linguistic characteristics of shill reviews. Electron Commer Res Appl 13(2):69–78. https://doi.org/10.1016/j.elerap.2013.10.002
24. Heydari A, Tavakoli M, Salim N (2016) Detection of fake opinions using time series. Expert Syst Appl 58:83–92. https://doi.org/10.1016/j.eswa.2016.03.020
25. Sharma K, Lin KI (2013) Review spam detector with rating consistency check. In: Proceedings of the 51st ACM southeast conference on—ACMSE '13, Savannah, Georgia, p 1. https://doi.org/10.1145/2498328.2500083
26. Saumya S, Singh JP (2018) Detection of spam reviews: a sentiment analysis approach. CSI Trans ICT 6(2):137–148. https://doi.org/10.1007/s40012-018-0193-0
27. WordNet | A Lexical Database for English (2010). https://wordnet.princeton.edu/. Accessed 23 May 2020
28. Klein D, Manning CD (2003) Accurate unlexicalized parsing. In: Proceedings of the 41st annual meeting on association for computational linguistics—ACL '03, vol 1, pp 423–430. Sapporo, Japan. https://doi.org/10.3115/1075096.1075150
29. Baccianella S, Esuli A, Sebastiani F (2010) SENTIWORDNET 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources evaluation (LREC10), p 5
30. SPSS Software | IBM (2011) https://www.ibm.com/analytics/spss-statistics-software. Accessed 23 May 2020

# Big Data Link Stability-Based Path Observation for Network Security


**Nedumaran Arappali, Melaku Tamene Mekonnen, Wondatir Teka Tefera, B. Barani Sundaram, and P. Karthika**

**Abstract** Wireless ad hoc network [WANET] systems are creating multihop communication structure between the mobiles to transfer the data groups. The remarkable qualities of remote frameworks cause the correspondence to interface among the conflicting mobiles. To manage high convey ability and biological blocks, most physical directing shows will not believe the stable associations during pack communication, which prompts the elevated delay and bundle reducing in the master-mind. The proposed research work recommends a way perception support physical steering convention that specifies POPR for WANET. The anticipated guiding show merges the relative partition, course and mid-expand forwarder center point with transfer thickness to propel data toward the objective in order to recover physical sending among the connection point. Multiplication results illustrate the projected directing convention, which performs superior to the existing arrangements.

**Keywords** Physical routing protocol · Direction · Traffic density · Link stability · Network security

## 1 Introduction

Transformation past specialists think about wireless ad hoc networks (WANETs) that empower unavoidable accessibility among mobiles and do not rely upon expensive framework system [1]. Correspondence among mobiles and earlier establishment release a lot of different sorts of talented applications for explorers with drivers. The applications are ought to give security with reassurance and help drivers to

N. Arappali (✉) · M. T. Mekonnen · W. T. Tefera
Department of Electrical and Computer Engineering, Wollo University, Kombolcha Institute of Technology, Kombolcha, Ethiopia

B. Barani Sundaram
Department of Computer Science, College of Informatics, Bule Hora University, Bule Hora, Ethiopia

P. Karthika
Department of Computer Application, Kalasalingam Univerity, Viruthunagar, Tamilnadu, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_36

mastermind and talk with one another so as to keep away from any mishap, road turned parking lot, inconspicuous impediments, speed infringement, climate data, sight and sound administrations, and so on [2]. Despite the fact that being a subclass of portable specially appointed systems, a remote system has a few properties that recognize it from other impromptu systems. The most imperative differentiations are high compactness plan; fast varying and dynamic topology which direct to high framework fragment and the disconnectivity in sort out [3].

## 2 Literature Survey

Wide range of networking applications is depending on productive bundle directing to upgrade the security and becomes open to driving condition [4]. A wide scope of steering conventions has been proposed to adapt to meager and profoundly versatile remote systems that are extensively gathered into various sorts. Physical directing conventions will build up resuscitated enthusiasm for versatile and remote systems [5]. In physical steering, the bundle sending choices depend on location of through the neighbors with objective center point. These shows were roughly arranged in group radio frameworks or for adaptable frameworks, and cannot map truly to remote frameworks. One of the basic purposes for these wonders was the improvement of the mobiles held in ways and ways are allowable by the circumstance [6]. In light of the vehicle thickness of the ways, the steering conventions must use limit data to accomplish the versatility prerequisites in the system [7]. Therefore, the vehicle hubs forward the parcels with the assistance of neighborhood data that are provided closer by direct neighbors. This procedure prompts fewer control transparency in light camouflage of the vehicle center point information of various bits of framework [8]. The GPSR show was projected in 2000 for remote data compose. It operates the circumstance of the mobiles and objective center point during method of group sending [9, 10].

## 3 Problem Identification

These mobiles be outfitted and satellite-based global positioning system (GPS) so as to choose vehicle territory and support multihop correspondence in arrange.

- Each automobile in arrange chooses the circumstance of its nearby convergences by preloaded electronic street level maps.
- Every automobile is in like manner responsive of remote traffic during a fundamental movement framework for way traffic judgment and transfer sensors introduced next to the intersections.
- The reference point messages be traded to recognize the nearby nearness, position with bearing in organize.

- The Dedicated short-range communication (DSRC) average is utilized for correspondence.
- Greatest sending separation is unchanging.

# 4 Implementation

## *4.1 Routing Protocol*

A hub has a lot of a jump nearby hubs within transmission extend that checks the ideal sending hub with various techniques. This vehicle hub is moving arbitrarily and habitually changes their location. Every vehicle hub intermittently communicates the guide messages to identify the portability attributes and get the data of one another. Along these lines, the guide messages contain the sequence of current location, time, bearing, and velocity obtained starting GPS.

## *4.2 Projected Routing Convention Sending Model*

Since prior to delineating the proposed coordinating show, the key idea and methods of reasoning are rapidly depicted. The procedure additionally checks the facilitator hub, and if the organizer hub is accessible the convention begins the crossing point-based need activity. In the event that there is no facilitator hub accessible, at that point its determination checks partition, heading and middle region center to propel the pack. Figure 1 shows sending model, anywhere the source center needs to throw the package toward objective.

## *4.3 Routing Measurements Between Crossing Point*

### 4.3.1 Distance and Path

The division and course are considered as noteworthy parameters. Since in transmission extent of vehicle center point, there is a probability that two centers are close to each other or of course they are confined through a partition of most outrageous radio range. The shorter partitions of vehicle center points prompt high number of bounces and nearest centers can deliver higher interface in organize. The likelihood of connection disappointment builds due to high flag lessening of problematic remote channels.

Fringe decision and ludicrous transmission go center completion may have a higher likelihood of leaving the radio range and dropping the bundles.


**Fig. 1** (POPR) operation between intersections and at intersection

### 4.3.2 Result Midexpanse Node for Next Hop Node

The sender vehicle center point picks the sending skip, in which heading is neighboring to the source and objective directly lines which progresses a similar way to stable parcel sending.

$$\text{Distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

The vehicle center point's course is obliged by ways and, in directly interstate circumstances, mobiles are touching in the identical or then again backward bearing. In Fig. 2, when the source center point intends to course the data package toward the objective, it sends the group from its directional neighbor center toward the objective center point. The source hub figures the mid-separation next bounce as pursues. The hid mid-zone is generally called affiliation zone with various neighbors inside the Radius1 and other with the $\text{Radius}_2$. The hidden zone $M_S$ can be resolute as.

$$M_s = M_1 + M_2 \tag{2}$$

where

$$M_1 = \text{Radius}_1^2 \cdot a_2 - \frac{\text{Radius}_1^2 \cdot \sin(2a_1)}{2} \tag{3}$$

$$B_2 = \text{Radius}_1^2 \cdot \alpha_2 - \frac{\text{Radius}_1^2 \cdot \sin(2\alpha_1)}{2} \tag{4}$$

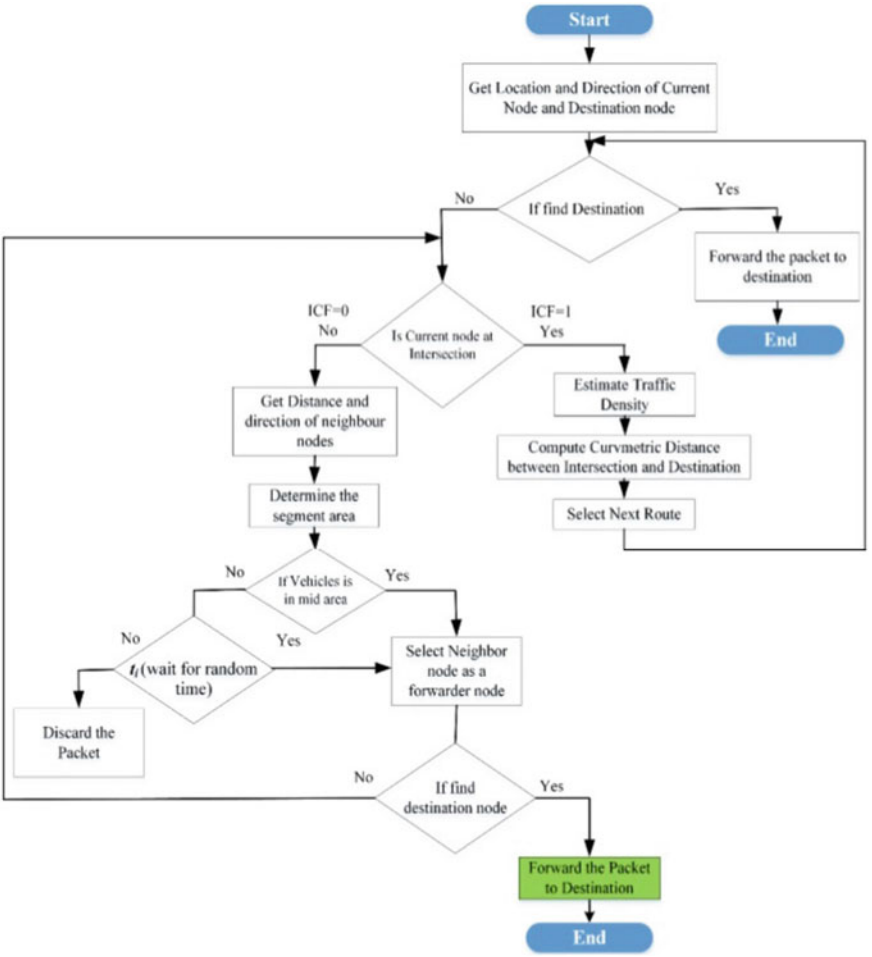



**Fig. 2** Architecture of POPR

The line among source center point and objective center point $90° = (\alpha_1 = 45°)$ is bisector of point; in this method, mid-region is $M_S$

$$A = \text{Radius}_1^2\left[\frac{\pi - 2}{4}\right] + \text{Radius}_2^2 \cdot \left[\alpha_2 - \frac{\sin(2\alpha_1)}{2}\right] \tag{5}$$

The mid-zone is a combine of two circular fragments $\text{Radius}_1$ and $\text{Radius}_2$ the inference of $\alpha_2$ rely upon the communication scope of source hub and the disconnection between source and goal hub. It ascertains the traffic thickness and heading

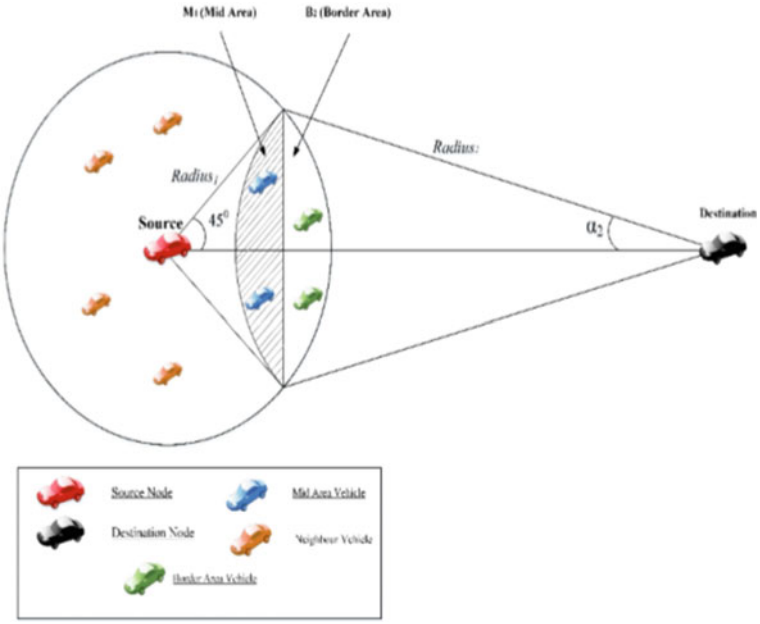



**Fig. 3** Mid-area with shaded line $M_s = (M_1 + B_2)$

toward the goal so as to choose next forwarder hub and course in Fig. 3. The greatness of path decision as the vehicle great ways from the transport center point to objective is as per the following

$$\mathrm{Dis}_{wv} = \max\left[\log\left[\frac{CD_{s,d}}{CD_{j,d}}\right], 0.1\right] \tag{6}$$

In condition (6), the $\mathrm{Dis}_{wv}$ shows the separation with weight worth and $CD_{s,d}$ shows the bend metric good ways from source hub to goal and $CD_{j,d}$ presents the curve metric good ways from current intersection to objective. In condition (7), the $\mathrm{Dir}_{wv}$(heading weight regard) with interface quality (LQ) factor is resolved between course of development vehicle and package transmission.

$$\mathrm{Dir}_{wv} = \left[\mathrm{LQ}\left(\overrightarrow{D}_n, \overrightarrow{D_{pt}}\right)\right] \tag{7}$$

When the vehicle gets a bundle, it discovers the weight score of each neighboring convergence dynamically traffic thickness ($TD_{wv}$) with typical number of mobiles ($N_{\mathrm{avg}}$) and reliable degree of system ($N_{\mathrm{con}}$) inside a telephone, as showed in condition (8).

$$TD_{wv} = [1 - D_c] + \left[\min\left[\frac{N_{\mathrm{avg}}}{N_{\mathrm{con}}}, 1\right]\right] \tag{8}$$

$\beta, \gamma$ and $\delta$ are utilized for weighting factors for separation, course, and traffic thickness, individually with $\beta+\gamma+\delta = 1$. Consequently, in view of the above examination, condition (9) characterizes the weighting score of the following convergence (Score (NI)).

$$\text{Score}(Nl) - \beta \times \max\left[\log\left[\frac{CD_{s,d}}{CD_{j,d}}, 0.1\right]\right] + \gamma \times \left[LQ\left[\overrightarrow{D_n}, \overrightarrow{D_{pt}}\right]\right] + \delta \times [1 - D_c] + \left[\min\left[\frac{N_{\text{avg}}}{N_{\text{con}}}, 1\right]\right] \tag{9}$$

The principle factor $\text{Dis}_{wv}$ is to check the partition to the objective in way length. The shorter detachment is favored toward the objective. The second factor $\text{Dir}_{wv}$ measures the vehicle heading with interface quality factor between developments vehicle and pack transmission. The last factor is the traffic thickness $TD_{wv}$ between the present and potential intersection point. In this tests, the vehicle transfer changes from 100 to 350 center points.

The Dijkstra calculation is utilized to locate the briefest ideal way with least weight. The DSR convention execution is not superior to CMGR and Geo SVR, in view of its way term expectation metric in Fig. 4.

The pattern in chart is enhanced for PDR contrasted and 150 hubs, since extra vehicle hubs in organize spread huge segments of the guide and broadcast extend. The POPR directing convention execution is enhanced within Fig. 5 for 250 hubs and a lot higher contrasted with meager system with 150 hubs.

At the point when the hub thickness expands, the briefest way along the ways turns out to be bound to have enough hubs and the PDR naturally increments.

The SDR execution is not superior to CMGR toward the beginning yet when the system will balance out, it is nearer to CMGR. In the preceding diagram, vehicle hubs are position to 350 shown in Fig. 6 and POPR execution is enhanced contrasted with

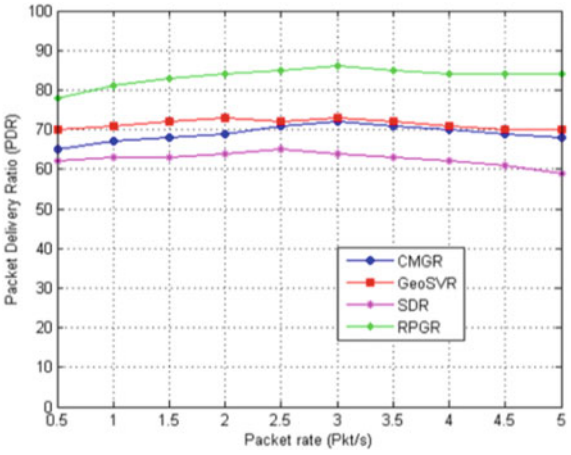



**Fig. 4** Average delivery ratio with 150 vehicle nodes

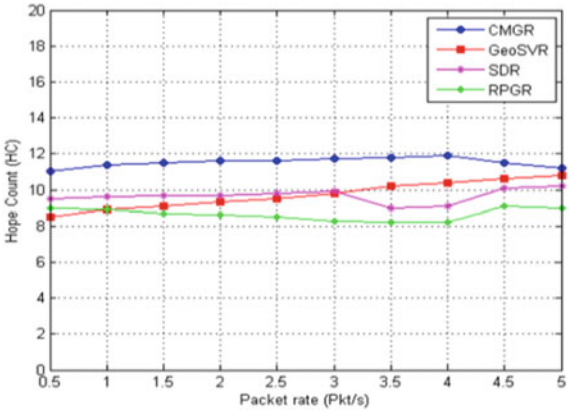



**Fig. 5** Average hop count with 250 vehicle nodes

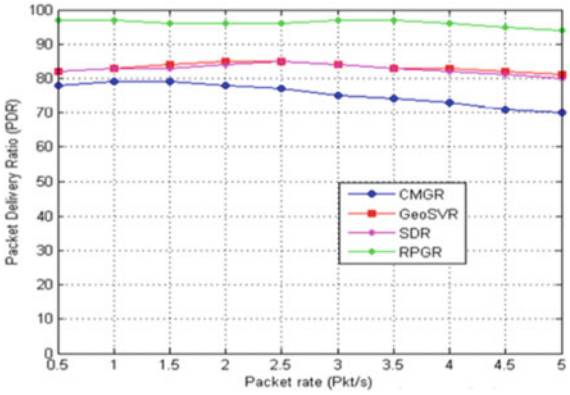



**Fig. 6** Average delivery ratio with 350 vehicle nodes

fewer thicknesses in organize. The POPR execution is better a result of its course and continuous traffic thickness measurements.

## 5 Result and Discussion

This corruption alludes to the bigger bundle size that prompts the higher data transfer capacity utilization and high immersion in remote channel. As appeared in Fig. 7, the PDR for all conventions diminishes attributable to colossal traffic load with enormous bundle size which at long last prompts high parcel misfortune.

The subsequent test in Fig. 8 plots the normal deferral of proposed POPR convention with other best in class directing conventions and dependent on the outcomes within the sight of various bundle measures, the POPR delay is not exactly the other. Then again, the SDR convention delay is superior to CMGR and Geo SVR.

**Fig. 7** Packet delivery ratio

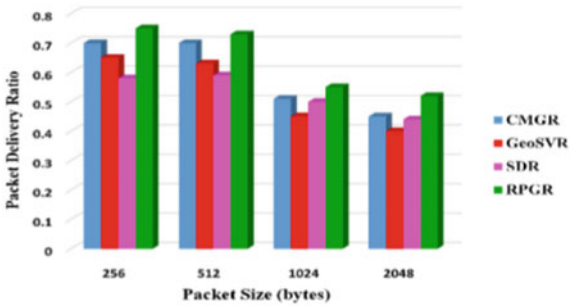
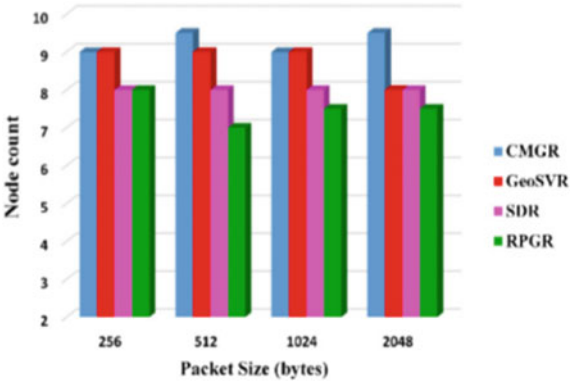



**Fig. 8** Node count




The CMGR convention bundle delay is high contrasted with different conventions attributable to the loss of some high idleness in information parcels.

The convention decides the mid-territory hub by checking the separation of neighbor hubs and heading toward the goal. This mid-territory determination offers bit of leeway to deal with high versatility and as often as possible changing topologies in remote systems.

## 6 Conclusion

This paper displays the POPR convention for bundle sending in WANET for urban-based condition. The POPR thinks about separation, course, and mid-locale hub to responsively choose the following crossing point. At whatever point the forwarder hub is at the crossing point, bend metric bearing and high traffic thickness are rechecked and forward the information toward goal in the system. The exhibition of the proposed steering convention has been dissected in reproduction with three existing CMGR, Geo SVR, and SDR directing conventions. The outcomes show

better execution of POPR as far as PDR, normal parcel defer and normal way length with various vehicle thickness and bundle size.

## References

1. Babu GR, Amudha V, Karthika P (2020) Architectures and protocols for next-generation cognitive networking. In: Singh KK, Cengiz K, Le D-N, Singh A (eds) Machine learning and cognitive computing for mobile communications and wireless networks (1st ed., pp 155–177). Scrivener Publishing Partnering with Wiley
2. Babu GR, Amudha V (2019) A survey on artificial intelligence techniques in cognitive radio networks. In: Abraham A, Dutta P, Mandal J, Bhattacharya A, Dutta S (eds) Emerging technologies in data mining and information security, vol 755. Advances in Intelligent Systems and Computing. Springer, Singapore, pp 99–110
3. Amudha V, Babu GR, Arunkumar K, Karunakaran A (2020) Machine learning based performance comparison of breast cancer detection using support vector machine. J Crit Rev. Accepted and in Press
4. Babu GR, Rajan DAJ, Maurya S (2019) Multi data users using novel privacy preserving protocol in cloud networks. In: Fourth international conference on information and communication technology for competitive strategies (ICTCS-2019) in association with CRC Press (Taylor and Francis Group), pp 883–890, Rajasthan, India
5. Babu GR, Obaidat M, Amudha V, Rajesh M, Sitharthan R (2020) Comparative analysis of distributive linear and nonlinear optimized spectrum sensing clustering techniques in cognitive radio networks. IET Networks, Accepted and in Press
6. Babu GR, Kumar MNS, Jayakumar R (2020) Dynamic exchange buffer switching and blocking control in wireless sensor networks. In: Mahapatra RP, Sharma R, Cengiz K (eds), Data security in internet of things based RFID and WSN systems applications, pp 155–177. CRC Press, Taylor & Francis Group
7. Nedumaran A, Babu RG (2020) MANET security routing protocols based on a machine learning technique (raspberry PIs). J Ambient Intell Humaniz Comput 11(7):1–15
8. Babu RG (2016) Helium's orbit internet of things (IoT) space. Int J Comput Sci Wirel Secur 03(02):123–124
9. Prabu S, Babu RG, Sengupta J, Perez de Prado R, Parameshachari BD (2020) A block bi-diagonalization-based pre-coding for indoor multiple-input-multiple-output-visible light communication system. Energies, MDPI AG 13(13):1–16
10. Krishnan R, Babu RG, Kaviya S, Kumar NP, Rahul C, Raman SS (2017) Software defined radio (SDR) Foundations, technology tradeoffs: a survey. In: IEEE international conference on power, control, signals & instrumentation engineering (ICPCSI' 17), pp 2677–2682. IEEE Press, Chennai, India

# Challenging Data Models and Data Confidentiality Through "Pay-As-You-Go" Approach Entity Resolution


**E. Laxmi Lydia, T. V. Madhusudhana Rao, K. Vijaya Kumar, A. Krishna Mohan, and Sumalatha Lingamgunta**

**Abstract** **Problem importance**: Predictive analytics seems to be an exceptionally complex and vital concern in domains like computer science, biology, agriculture, business, and national security. When big data applications were indeed accessible, highly efficient cooperation processes are often meaningful. Simultaneously time, new subjective norms originate when the high quantities of data will conveniently assert confidential data. This paper has reviewed two complementary huge issues: data integration and privacy, the ER "pay-as-you-go" approach (Whang et al. in IEEE Trans Knowl Data Eng 25(5):1111–1124 (2012) [1]) in which it explores how the developments of ER is maximized to short-term work. Stepwise ER problem (Whang and Molina in PVLDB 3(1):1326–1337 (2010) [2]) is not even a unique process; it is done concurrently by the better usage of information, schemes, and applications. Joint ER problem with multiple independent datasets are fixed in collaboration (Whang and Molina in ICDE (2012) [3]) and the problem of ER with inconsistencies (Whang et al. in VLDB J 18(6):1261–1277 (2009) [4]). To overcome the research gap in the existing system, the proposed research work addresses an entity resolution (ER) problem that tends to address the records in databases referring to a certain complex real-time entity.

**Keywords** Data analytics · Data availability · Privacy · Entity resolution

E. Laxmi Lydia (✉) · T. V. Madhusudhana Rao
Department of Computer Science and Engineering, Vignan's Institute of Information Technology, Visakhapatnam, India

K. Vijaya Kumar
Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India

A. Krishna Mohan · S. Lingamgunta
Department of Computer Science and Engineering, JNTUK, Kakinada, Andhra Pradesh, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_37

## 1 Introduction

Application data integration priorities for real world suggest methods efficiently and produce records for optimized ER outputs while reducing data size and to research how ER implementations can use generated hints. Maintaining revised ER outcomes as the ER framework used in record analysis changes frequently.

A functional framework of modular resolution in which developed ER algorithms, built over a certain record form can be embedded into and used with many ER algorithms. Prohibit ER solution irrationality using ER negative laws

Data privacy policy seeks to establish an information security model that considers the loss of personal privacy issues on an ongoing scale ranging from zero to one, provides effective computational leakage algorithms, assesses the efficiency of the information and its scalability to control information leakage, and suggests deception methods for agency resolution.

Flexible ER techniques and different ER features are also introduced. ER is often considered as a black-box process and includes basic techniques for operations. The issues of handling data corruption are initiated, in which sensitive information bits should be shielded by ER in order to prevent unnecessary loss. When a number of retailers, healthcare professionals, staff, networking media and so forth are more open to our personal data, there is an average increase that an attacker will link the dots and put our metadata together, resulting in much more privacy loss. This paper suggests a method to estimate the leakage of information and use "disinformation" to prevent leakage of information.

To fill the gap of existing system in this paper, an entity resolution (ER) problem that addresses records in databases referring to a certain complex real-world entity. Related to the recent digital revolution, ER seems to have become a challenge in a variety of technologies. We also suggested the use of modular ER methods and tools for huge datasets. We apply terms of strategy since they anticipate either its complete ER phase as a black-box deployment and can therefore be included in a large number of ER applications, or perhaps the refer fusion mechanisms. We have suggested a paid-by-go method for ER, in which we focus on making optimum progression despite the limitations in basics (e.g., function and runtime).

## 2 Literature Status

**Andrew K. McCallum (2000), Rohan Baxter (2003)**: The blocking strategies aim to increase the optimal ER operating time in which documents can be separated into inconsistent blocks and indeed the blocks are set one by one [5, 6].

**Omar Benjelloun (2009), Christopher D. Manning (2008)**: Contrasting documents and deciding whether they are related from the same entity is part of the precision of the entity. Much of the job is part of one of the ERs that match-based clustering is

taken into account [7]. The ER work involves elaborating ER's precision or latency efficiency, but typically assumes a predetermined logic for records resolution. As far as we recognize, our task will be the one to take into account the ER problem until the alternative logic modifications itself [8, 9].

**Steven E. Whang (2011)**: ER-based actions to estimate the quantity of confidential information published to the public [10]

**Steven Euijong Whang (2012)**: This paper examines how to significantly increase innovations of ER utilizing "hinting" in a constricted volume of work to apprise everyone about records probably belong to the same real-world organization everywhere. A hint is worthy of consideration with different configurations (for example, an aggregation of records relying on certain chances of overlapping), and ER would use the data as a recommendation to evaluate records first. A class of procedures are implemented for the powerful construction of hints and strategies for the maximization of compatible records with a minimal workload. The proposed significant value as a pay-as-you-go approach is correlated and displayed to implement ER without the use of hints with virtual datasets [3].

**Steven Euijong Whang, Hector Garcia-Molina (2013)**: Their methods can be combined with the disinformation issue. It has been supposed that while an "agent" seems to have certain sensitive information to obtain from the "adversary." For instance, a cameras corporation might secretly design a latest camera prototype and a customer (the opponent) might also want to understand the full specifications in ahead of time. The purpose of this agent will be "dilute" what the opponent knows by disseminating false information. The opponent is structured as an entity resolution (ER) mechanism, which compiles resources displayed. They ratify the question of determining the most beneficial disinformation with a small budget for disinformation and performance. They officialise the challenge of disinformation with a restricted resource and recommend optimization techniques to address the situation with the utmost importance. After this, they scrutinize the rigidity of emerging ER algorithms with our methodologies for disinformation scheduling on ultimate and synthetic information. The disinformation methods may usually be enough to assess ER durability [11].

**Gagan Aggarwal (2004)**: The P4P policy is designed to cover the unauthorized use of private information that is now being revealed to an opponent. For relevant information, mechanisms of multipurpose are formulated to maintain data control system [12].

**Rui Xu (2005)**: Clustering strategies streamlined toward noise were thoroughly examined in the past. With in face of excessive noise, this kind of work suggests learning algorithms that obtain the exact clusters. Conversely, the aim is to deliberately distract the ER technique even more than necessary for the target entity [13].

**Arvind Arasu (2006)**: Some activities suggest efficient combinations of similarity. Our pay-as-you-go methods strengthen blocking through just using the ordering of record pairs to obtain the maximum intermediate ER performance. [14].

**E. Laxmi Lydia et al. (2016)**: K-means is one of the most basic and unsupervised learning techniques that resolve any well-known clustering problem. Suggest a cluster K number for the quick and fast classification of grid processor. Since K-means clustering is centered on cluster centroids, there is no guarantee of an excellent solution. The framework then uses the clustering of K-centroids [15].

**E. Laxmi Lydia et al. (2016)**: With author's ongoing research on the disparateness of cluster system, resource type, processing speed, and memory features are developed. The device must form a cluster with the K-centroid clustering to prevent a scheduled interruption. The node transfers to the cluster [16] are based on higher key objectives.

## 3 Detailed Methodologies of the Research

**Existing System**: This paper has investigated two distinctive complicated issues: data integration and data confidentiality, the ER "pay-as-you-go" approach [1] in which it examines where to achieve maximum growth of ER with a small workload. Sequential ER problem [2] ER is not really a unique process, and it is carefully examined by the proper understanding of data, schemes, and applications. Joint ER problem for which multiple individual datasets are handled in combination [3] and the incoherence phenomenon of ER [4].

**New Area to Fill the Gap of Existing System**: This project will instead address the issue of entity resolution (ER), where database records are identified that further correspond to almost the same real-world entity. Due to the recent data explosion, ER has become a challenge in a multitude of scenarios. Further, it also suggests the use of modular ER techniques for complex data. The overall strategies are applied since they recognize whether another entire ER process as either a black-box operation and therefore can be used in a vast array of ER application domains, or the contest and fusion functions. A pay-as-you-go method is suggested for ER to make optimum advancement considering the problems in resources (e.g., function, runtime).

The extensible ER methodologies and new ER features, which were not previously practiced are utilized here. ER is often interpreted as black-box process and has simple basic applications that will be used. Besides that we address the issue of information leakage management, where you have to work to prove that sensitive info bits are not played with by ER and defend toward data privacy losses. When our personal details become more accessible to a number of vendors, distributors of healthcare, workers, Web media, and perhaps more, an attacker seems to be more likely to "connect the dots" to our information and to bring it together, contributing to an

increased risk of confidentiality. A quantitative measure is asserted for leakage and to use "disinformation" as a device to incorporate disclosure of sensitive information.

### *3.1 The Following Portrays the Proposed Model in Detail*

Also because of very massive datasets and complex record comparative analysis, an ER approach is generally cost-effective. The processing of platforms such as social media, for illustration, will generate thousands or even millions of data which has to be remedied. It can be expensive to check each pair of documents in order to progress their "similarity," as distinct fields are contrasted, invoking substantive application logic. Around the same time, running ER in a quite short time can be very critical. The "pay-as-you-go" model has been utilized to perform entity precision, which slowly yields extracted features by resolving them. As observed, all documents belonging to the same real-world individuals cannot be classified by preliminary results. Our ultimate objective seems to be to gather even more of the eventual outcome as possible, as quickly as necessary. It is significant to mention which our current job is qualitative by existence. Suggested hints are identified as heuristics. Further, our overall objective is to attain a collaborative platform for hints and to assessing the potential improvements.

The following Fig. 1 demonstrating the flowchart of methodology steps in brief.

### *3.2 The Methodology Adapted*

- Pay-as-you-go ER is formally accepted model attempt to optimize the temporary ER consequence. Our blocking techniques depend on the regular ER control flow.
- Various levels of hints are suggested:
  - The most detailed (and yet compact) form of hint: the sorted list of record pairs.
  - A hint that is relatively informative and portable: A hierarchy of partitions
  - The most compressed form of hint (and yet least informative): Sorted records
- The techniques are proposed with successful hints generation at each hint form and scrutinize whether ER iterations can optimize ER efficiency whereas limiting record number correlations.
- Our approach has been enlarged to include multiple hints.
- It will quantitavely consider what ER can yield results quickly with the application of tips. We use valid comparison consumer details from Yahoo! Shopping and hotel statistics from Yahoo! Travel. Our findings indicate situations in which suggestions boost the ER filtering such that most corresponding records can be found in a proportion of the overall runtime.

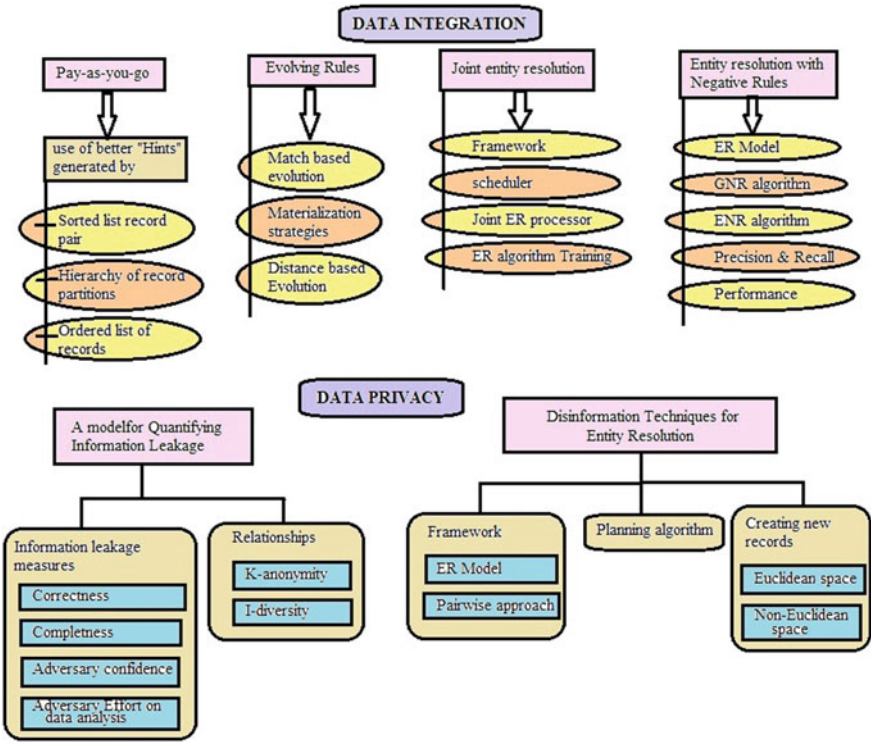



**Fig. 1** Flowchart of the methodology

### 3.2.1 System: Architecture and ER Design

This section describes the proposed pay-as-you-go ER system in this module. Initially, a general business determined model is proposed, followed by an overview of where pay-as-you-go works.

A database collection R, which represents real-world entities, is used as an EE encoder E. The ER performance is an input partition which mostly contains information representing a certain real-world object separately. Case: output $P = \{\{e1, e3\}, \{e2\}, \{e4, e5, e6\}\}$ means that e1 and e3 of records are an individual entity, and e2 is an individual entity, etc. Since it sometimes requires to run ER on the results of a previous resolution, developers have defined this as a fraction. That record in its own folder directly, e.g., {{e1}, {e2},} e4, {e5}, {e6}}. This sense that report is a single partition. The ER product of $R$ at time t is revealed that $E(R)[t]$.

### 3.2.2 Description for Pay-As-You-Go Model

The pairs of candidates are concretely organized with pay-as-you-go template by the probability of a matchup. So, the ER approach identifies the record with the more probable pairs instead. The objective is obviously, since the requirement is unpredictable to evaluate the couple sequence very effectively.

To display this, six records have been put in two blocks: one block holds the records e1, e2, and e3, and the second block holds the records e4, e5, and e6. The candidate pairs are indicated by {e1–e2, e1–e3, e2–e3, e4–e5 ….}. A conventional ER algorithm will indeed analyze between pairs, perhaps in some random order regarding most of these pairs during the first block as well as the second block pair. However, the most feasible pair is equated and assumes as d5-r6, with pay-as-you-go. Then, it probably correlates the very next, say e2-e3. Fortunately, every block is authorized autonomously if only yet another block at such a time fits into the memory. Specifically, first of all the first block pairs are equated by the chance of a downward match, and then the second block seems to be the same. In any case, the solution is to determine compatible pairs more quickly than from randomly analyzing the candidate pairs. The ER automated system could then build an output portion which approaches the overall product more rapidly. (To measure performance continuously and to treat candidate pairs as a consequence of increasing probability, it is not like all ER methods can, as stated previously.)

The first prerequisite is our key objective of providing early ER outcomes of higher quality. The second prerequisite assures that the pay-as-you-go defines the findings preceding the ER of almost the same accuracy as the outcome without any hint. Blocking techniques will in response give an estimated ER result for which output has slipped. The auxiliary data format suggestions are used to effectively produce candidate pairs in (equivalent) sequence of probability. They address three kinds of clues, as shown in Fig. 2.

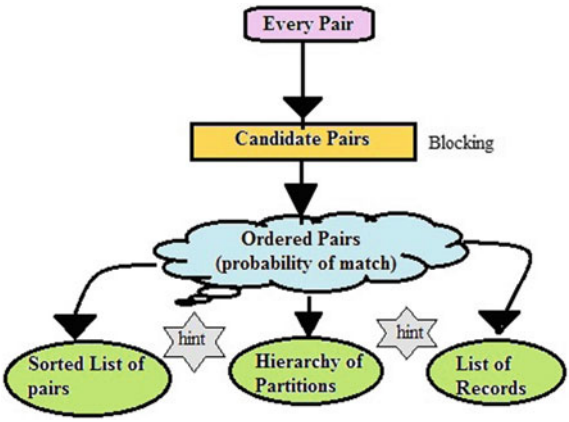



**Fig. 2** Implemented framework for pay-as-you-go model

### *3.3 Implementation of Pay-AS-You-Go Model Through Ordered Pairs with Sorted List*

This section discusses a hint that provides a list of record pairs that mostly fit pairs. It has been claimed that both the distance and match function would be used for the ER algorithm. The distance function $d(r, s)$ has been used to quantify the difference among both $r$ and $s$ records: as the distance is shorter, r and s display the same truth. A match function $m(r, s)$ is evaluated as valid if $r$ and s are assumed to be another object in the real life. It must be noted that may be a match feature may use a remote. This might be the form of "if $d(r, s) < T$ and subsequently valid" when $T$ is a threshold, for starters. It can also be assumed that $e(r, s)$ is much affordable than $m(r, s)$ and $d(r, s)$ for calibration. The ER technology requires an equivalent function, the relatively small $e(r, s)$ valuation, the greater the likelihood of the $m(r, s)$ being counted as valid. The sequence of all record pairs, procured by the through value, is another conceptual suggestion. The collection cannot in action be produced completely and precisely. For comparative purposes, beyond a certain series of pairs and even after the predictions reach a certain threshold the category may be distorted. As you can see, the pair seems to be another substitute: The ER algorithm will query the next pair upon this list, as well as the pair is determined at that juncture. Then, an $O(N_2)$ complexity can be avoided when the hint is generated. The proposed research work will converse over how to produce a pair list indication and using a pair list more with ER algorithm.

### *3.4 Generation*

Many ways are suggested to develop a portion hierarchy adequately. Within the next segment, the hints are created with which the application predictions are more focused on the sorted records. Then, this section will explain how memory hierarchies can sometimes be produced with hash (also implementation estimates, but not evaluation predictions) and affordable distance functions.

#### 3.4.1 Using Assumptions of Applications

For certain instances, an application-specific prediction function may be built which is inexpensive to measure. For comparison purposes, if indeed the remote function calculates a gap among records, then zip code is used to calculate the distance: if two records use the same zipper code, then it will be conveyed that they are close; otherwise, we will say individuals are far. The approximation can hardly recognize one or two attributes, yet the most impressive, unless the distance function calculates and merges the similarity among most of those of the record's features.

In order to develop the hint, $e(r, s)$ are calculated including all record couples but every pair can be inserted and estimated in a heap data structure. Once all pairs have been added, where it might delete all couples by increasing estimates, because it is highly required to complete the list. However, developers could even delete entries until we hit a threshold distance just a couple of pairs just before ER algorithm begins to demand challenging pairs.

For many other situations, the projections project over a single dimension to distances, though in that case a large reduction throughout the data amount in the heap will take place. For instance, $e(r, s)$ has been the difference in quality of records. In this case, the records could be sorted by value. (Notify them a certain records near as possible in price seem to be likely to accept) Then, on the cost dimensions (and consequent cost difference), we insert into the heap with each record, its nearest neighbor. The record r is obtained from the heap with the closest neighbor to get the smallest estimate set. R's next nearby neighbor is subsequently searched (in the sorted list), and the new assessment is used to integrate r into the heap. Throughout this case, the maximum capacity is analogous to, the record number. From the other hand, the maximum capacity order whenever all pairs of records will remain on a heap as $O(|R|_2)$.

### 3.4.2 Implementation of Pay-As-You-Go Model Through Hierarchy of Record Partitions

We suggest the hierarchy of partitions as a potential structure for hints in this statement. A result hierarchy provides details on possibly complete analysis in the context of scoreboards with various granularity rates where each score reflects a "potential world" of an ER outcome. The lower-level portion is the finest clustering of the raw data. Less rough grain with bigger clusters is greater collections in the Hierarchy. Namely folders in order of granularity, during which grosser cuts are extremely high up in the hierarchy, without ever storage of arbitrarily defined partitions.

**Algorithm 1** Developing a hierarchy of partitions from sorted records.

Input: Two lists such assorted records and Thresholds were considered Sorted = [r1, r2,….] and T = [T1,…..TL]

Output: Aim is to achieve a hint H = {S1,…, SL}

Process:

3: Start with S1,…., SL partitions

4: Perform for r ∈ Sorted

5: Perform for Tj ∈ sorted

6: check for the condition if r.prev.exists() ˆ KeyDistance(r.key, r.prev.key) ≤ Tj then

7: Extend r to Cluster Sj

8: else

9: Build a new Sj cluster with r

10: return {S1, …., SL}

The Algorithm 1 demonstrates when to create a hint H partition hierarchy employing various T1 and TL thresholds, regarding the key value length of partitioning information. For example, suppose a 3-record list of records [joy, joyful, joyous] (the records are listed and sorted by names) is included. (The thresholds are predefined focused on level L count in H) Believe that set 2 thresholds T1 = 1 and T2 = 2, and use the editing interval. For the purpose of highlighting the differences between the records, thus eliminating the right to shift the series to the next. Second, joy reads the algorithm 1 and applies it to the existing S1 and S2 cluster (Step 9). The word joyful will be added in the first cluster in S2 (Step 7), because the edit distance does not exceed T2. The last record of joyous is 4, which satisfies all thresholds. Change difference only with previous estimate joyful is 4. A conjugating for both S1 and S2 is therefore established with joyous. Thus, there are two partitions in the subsequent hint:

S1 = ffjoyg, fjoyfulg, fjoyousgg and S2 = {{joy, joyful},{joyous}.

The results below display first algorithm consistency.

***Proposition 2.3.1*** A relevant hint is given by the first algorithm.

***Proof*** A partition S of higher rates is often grosser than even a partition S0, whereas an increased threshold has been used in the sorted list to break data. S0_S, therefore, since partitions of the highest level is often more delicate than partitions of the lower level, H is the description hint. Since the sorted input has now been sorted, Algorithm 1 is used to iterate all records in sorted in O(L) time and iterates across all threshing levels within each record.

## 4 Requirements for Experimental Analysis

This paper will experiment with a retail comparison dataset from Yahoo! that maintains (shopping and hotel) data. This will test the SN, HCB, HCBR, ME, HCDS, and HCDC algorithms for ERs. Algorithms were performed testing using the 2.4 GHz Intel (R) Core 2, 4 GB RAM processor in Java.

The detailed software information's are as following:

| Language | Java programming |
|---|---|
| Used system | checked only in GNU/Linux |
| IDE supporter | NetBeans IDE 6.0.1 |
| UML software documentation | Umbrello UML Modeler 2.0.3 |

As Hadoop and Mahout function locally with Java, accessibility between components generated by Java would be easy. Considering this reality, Java was picked as the programming language. Umbrello UML Modeler has a basic yet effective arrangement of demonstrating instruments, because of which it was utilized for UML documentation. NetBeans IDE was picked as formative IDE accounting to its rich arrangement of elements and simple GUI Builder tool.

## *4.1 Real Data*

Yahoo! Shopping issued the comparative assessment shopping dataset that has been used here and includes millions of records that turn up consistently from divergent online stores and have to be addressed before customer queries are replied. Almost every database includes different attributes, except the title, price, and object type. The plan checked a small portion of 3000 shopping documents with the string "iPod," and a random subsystem of 1 million purchasing files. This was also checked with a hotel dataset given by Yahoo! Where decades of thousands of customers accumulate and must be processed prior to users travel from various sources (e.g., Orbitz.com) being seen. The 3000 accommodation information will be checked in the USA, which is a randomized subset. The 3 K shopping and hotel statistics are memory capable, but the 1 million shopping datasets are memory capable and have to be reserved from the server.

## *4.2 Guidelines of Match*

The match guidelines in our experimental work are summarized in Table 1. The sort attribute specifies yet if the match requirements are Boolean or RM rules. The details of the field indicate the source of data: shopping or hotel information. The column for the match requirements shows the match regulations apply. Boolean rules on shopping and hotel datasets are specified throughout the first two rows. $B_1^S$ relates two retail documents "names and ranges, while $B_2^S$ relates the retail records" titles and costs. $B_1^H$ relates the state, city, zip codes, and names of the hotel documentation for the information on the data analysis of the hotel. The $B_2^H$ rule contrasts two hotel files with states, towns, postal code, and home locations. Later, $D_1^M$ measures the

Jaro distance between adjacent shopping item titles while $D_2^D$ random modifications distance from $D_1^S$ to fulfill the demands of 5 percent. The Jaro distance is given to the [0, 1] quantity of the shopping record and appears to give larger levels to close databases. The Jaro value is calculated to a maximum ratio of 5 percent with each of the two rows. Again $D_1^H$ accumulates the gap from Jaro for hotel data from the descriptions of two records and indeed the equality including its two applications weighted at 0.05 from the regions. Nowadays, the equality distance increases to bring back one if two values are identical and 0 and if they are not identical. The $D_2^H$ rule perfectly illustrates the Jaro distance between certain names of the zip codes weighed by 0.05 with the appropriate intervals between them. As a matter of fact, at majority the 0.05 ER and rule evolution methodology that affects the altitude from $D_1^H$. This paper is experimental with the development of SN, HCB, HCBR, ME, HCDS and HCDC. Following Table 2 shows each ER implementation and the rule evolution mechanism which has been used. (This does not undergo rule progression rather than using the join ER model, however, emphasizes upon on clustering ER model) The remote advanced analytics HCDS and HCDC end with a fixed distance of around the threshold of 0.95 among both clusters (remember relatively close accounts have relatively high Jaro + Distance inclusiveness). Whereas ME and HCDC models do not comply with RM, Algorithm 7 can indeed be implemented to develop quality and limited loss of consistency new ER outcomes. Confirm that, while ME is GI, Algorithm 3 is not reliable, as ME derives and resorts all documents in the current

**Table 1** Match rules

| Type | Data | Match rules |
|---|---|---|
| Boolean | Shopping | $B_1^S : p_{ti} \wedge p_{ca}$<br>$B_2^S : p_{ti} \wedge p_{pr}$ |
| Boolean | Hotel | $B_1^H : p_{st} \wedge p_{ci} \wedge p_{zi} \wedge p_{na}$<br>$B_2^H : p_{st} \wedge p_{ci} \wedge p_{zi} \wedge p_{sa}$ |
| Distance | Shopping | $D_1^S$ : Jaro$_{ti}$<br>$D_1^S$ : Jaro$_{ti}$ with random shifts in less than 5% |
| Distance | Hotel | $D_1^H$ : Jaro$_{na} + 0.05 \times$ Equals$_{ci}$<br>$D_2^H$ : Jaro$_{na} + 0.05 \times$ Equals$_{zi}$ |

**Table 2** Referenced implementations for ER and rule evolution

| ER algorithm | Algorithm for rule evolution |
|---|---|
| SN | Algorithm for SN |
| $HC_B$ | Algorithm 3 |
| $HC_{BR}$ | Algorithm 7 |
| ME | Algorithm 7 |
| $HC_{DS}$ | Algorithm 7 |
| $HC_{DC}$ | Algorithm 7 |

partition Pi (not through any of the Pi clusters). Algorithm 7 will be used for the distance-based clustering pattern, including both HCDS also HCDC computational methods.

## 5 Conclusion

The paper reports primarily concerns about data processing to tackle the issues of computer science, and perhaps other relevant fields besides the finance and medicine. Then, the proposed research addressed the entity resolution question (ER), which mostly recognizes database records belonging to a certain real-world object, in the paper implemented. The pay-as-you-go method is suggested for ER, in which attempts to achieve and produce the majority and given a constraints (e.g., job, runtime). In order to stop sensitive pieces of documents that are being resolved by ER, the issue of handling the leakage of information is incorporated to protect toward loss of privacy. For more of our personal details disclosed to a number of vendors, care professionals, supervisors, and other networking media, there is a greater risk which an offender will "mark the dots." The estimation has been formulated to evaluate the leakages while using "disinformation" as a platform to constitute leakage.

## References

1. Whang SE, Marmaros D, Molina HG (2012) Pay-As-You-Go entity resolution. IEEE Trans Knowl, Data Eng
2. Whang SE, Molina HG (2010) Entity resolution with evolving rules. PVLDB 3(1):1326–1337
3. Whang SE, Molina HG (2012) Joint entity resolution. In: ICDE
4. Whang SE, Benjelloun O, Molina HG (2009) Generic entity resolution with negative rules. VLDB J. 18(6):1261–1277
5. Baxter R, Christen P, Churches T (2003) A comparison of fast blocking methods for record linkage. In Proceeding of ACM SIGKDD. Workshop on data cleaning, record linkage, and object consolidation
6. McCallum AK, Nigam K, Ungar L (2000) Efficient clustering of highdimensional data sets with application to reference matching. In: Proceeding of KDD, pp 169–178, Boston, MA
7. Bhattacharya I, Getoor L (2004) Iterative record linkage for cleaning and integration. In: DMKD
8. Manning CD, Raghavan P, Schtze H (2008) Introduction to information retrieval. Cambridge University Press, New York, NY, USA
9. Benjelloun O, Molina HG, Menestrina D, Su Q, Whang SE, Widom J (2009) Swoosh: a generic approach to entity resolution. VLDB J 18(1):255–276
10. Whang SE, Molina HG (2011) A model for quantifying information leakage. Technical report, Stanford University
11. Whang SE, Molina HG (2013) Disinformation techniques for entity resolution. Stanford University, Canada. ISBN 978-1-4503-2263-8
12. Aggarwal G, Bawa M, Ganesan P, Molina HG, Kenthapadi K, Mishra N, Motwani R, Srivastava U, Thomas D, Widom J, Xu Y (2004) Vision paper: enabling privacy for the paranoids. In: VLDB, pp 708–719

13. Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans Neural Netw 16(3):645–678
14. Arasu A, Ganti V, Kaushik R (2006) Efficient exact set-similarity joins. In: VLDB, pp 918–929
15. Lydia EL, BenSwarup M, Narsimham C (2015) A disparateness-aware scheduling using K-centroids clustering and PSO techniques in hadoop cluster
16. Laxmi CHVTEV, Somasundaram K (2015) 2HARS: heterogeneity-aware resource scheduling in grid environment using K-centroids clustering And PSO techniques

# Preserving and Scrambling of Health Records with Multiple Owner Access Using Enhanced Break-Glass Algorithm


**Kshitij U. Pimple and Nilima M. Dongre**

**Abstract** Cloud computing servers gives a stage for the clients to remotely store information and offer the information to everyone. Healthcare record management system (HRMS) has been developed as a tolerant-driven model for health-related data exchange. Classification of the common information, i.e., visit details, report details, etc., remains as a serious issue when patients utilize public cloud servers since it might be viewed by everyone. To guarantee the patient's control over admittance to their own uncommon well-being records, it is a promising technique to encode the reports before redistributing and pick up the gain power to that data. Security introduction, adaptability in key association, and flexible access experience have remained the most essential difficulties toward accomplishing fine-grained, cryptographically endorsed data get the opportunity to control. This research work proposes an arrangement of systems for information get to control to healthcare record management system (HRMS) put away in outsider servers with classification of data into public and private data. In HRMS for accomplishing smooth and adaptable information get the chance to control, attribute-based encryption (ABE) methodologies are utilized to encode each patient's HRMS record for getting to. While providing secure data outsourcing, the main concentration is the multiple owners of data. On account of this framework significantly diminishes the key administration unpredictability for information owners and clients, HRMS ensured a high level of patient security. With respect to emergency situations, a break-glass extraction method is introduced here to access the data in emergency situations with authorized keys.

**Keywords** Cloud computing · Cryptography · Attribute-based encryption

K. U. Pimple (✉) · N. M. Dongre
Department of Information Technology, Ramrao Adik Institute of Technology, Nerul Navi, Mumbai, India

N. M. Dongre
e-mail: nilima.dongre@rait.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_38

## 1 Introduction

The Internet has become very vastly as of late, and it offers different abilities that could bolster physicians to perform their responsibilities from multiple points of view. Cloud computing (Fig. 1) is the conveyance of processing administrations including servers, stockpiling, databases, organizing, programming, investigation, and insight—over the Internet ("the cloud") to offer quicker development, adaptable assets, and economies of scale.

With the advancement of data and restorative innovation, medical information has been changed from manual written records into e-records into electronic therapeutic records which are extensively used. Nowadays, programming structures have created from the individual customer's neighborhood equipment to a focal server that works from a remote region. A health record is nothing but record of a person user's health data put away in PC and can oversee, track, and take an interest in possess medicinal services. The benefits of data-driven healthcare systems, in any case, accompany a value, an exceptional exertion and effort to secure patients' protection without trading off the utility of the information and related human services administrations. Like in training and interpersonal interaction, securing privacy in medicinal services frameworks is critical and testing as a result of the affectability of the related information and the multifaceted introduction of the healthcare systems.

Unapproved presentation of the touchy medicinal services information can have profound established social, financial, and well-being related outcomes to the patients. Masters, orderlies, lab experts, emergency responders, and cure handlers comprise the healthcare systems and approach patients' data. Additionally, specialists and other model shoppers require patients' information and the created models for their logical investigations and business purposes, including some possibly against patients' best preferences. Furthermore, the sheer measure of information and related multifaceted nature of investigation require human services frameworks to utilize outsider cloud foundation suppliers for enormous scale parallel taking care of and


**Fig. 1** Cloud computing [23]

data stockpiling. These exposures cannot be discarded because of the basic jobs the taking an interest substances play. Ensuring security requires guaranteeing that individuals keep up the legitimately to control what information is accumulated about them, who takes care of it, who uses it, how it is used, and what reason it is used for.

## *1.1 Data Privacy and Security*

Data privacy can be defined as:

(1) Untraceability: Making it problematic for an adversary to recognize that a comparable subject played out a given course of action of exercises.
(2) Unlinkability: Concealing information about the association between any things, for instance, subjects, messages, activities, and so forth.
(3) Unobservability: Concealing how a message was sent (rather than covering the identity of the sender of message)
(4) Anonymity: Concealing information who played out a given action or who is delineated by a given dataset.
(5) Pseudonymity: Using pen names of utilizing genuine identifiers.

Giving security requires counteracting access to information or unique articles by unapproved customers [4, 5], similarly as guaranteeing against unapproved changes or pummeling of customers' information.

The commendable significance of security levels with it, with characterization, decency, and openness, called (by its contraction) the CIA gathering of three. The ISO 7498-24 standard stretches out the security definition to the going with necessities that can use the gathering CIA-AANN:

(1) Confidentiality: Data is not made open or uncovered to unapproved individuals, substances, or techniques.
(2) Integrity: Information has not been changed or obliterated in an unapproved way.
(3) Availability: A system is operational and handy at a given moment (it isn't down).
(4) Access control: Clients get to simply those advantages and organizations that they are equipped for access, and qualified customers are not denied access to organizations that they sincerely plan to get.
(5) Authentication: The confirmation that a component is the one declared, and the wellspring of got data is as ensured.
(6) Nonrepudiation: The senders/recipients of messages cannot deny that they actually sent/got the messages.
(7) Notarization: The enrolment of data with a trusted in outcast that ensures the precision of data traits, for instance, substance, origination, and creation time. Protection and security are specific. Regardless, in human administrations Fifield, the association between those two thoughts is data affirmation. The

basic association exhibits among security and protection in the beneath figure; that is, confidentiality is situated in the convergence of protection and security.

## *1.2 Data Security in Health Care*

Healthcare record management system (HRMS) has created as a tolerant driven model of well-being data trade. Commonly displaying HRMS organizations for each individual, there are various security and safety risks that may block its huge selection. The focal concern is about whether or not the patients could truly control the sharing of their fragile prosperity realities, particularly when they are made sure about on a public worker which people may not completely trust. From one perspective, regardless of the way that there exist human administrations rules, for instance, HIPAA which is starting late changed to join colleagues cloud suppliers are commonly not verified substances. Obviously, because of the high assessment of the delicate individual health information, the untouchable storing workers are regularly the objectives of different poisonous practices which may impel introduction of individual health information. As a lofty occasion, a Department of Veterans Affairs information base containing fragile individual prosperity information of 26.5 million military veterans, including their administration oversaw investment funds nos. also, clinical issues was burglarized by an agent who took the data outside without endorsement. To guarantee lenient-driven security control over their own HRMS, it is essential to have fine-grained information get the chance to control parts that work with semi-trusted in workers. Therefore, another framework is recommended that guarantees the security of HRMS that ought to be finished utilizing ABE algorithm. HRMS association enables a patient to administer and control her own extraordinary thriving data in a lone spot through the web, which has made the cutoff, recuperation, and sharing of the remedial information continuously possible. The principle objective is to verify individual health records utilizing multi-expert ABE. That is various experts can get to the HRMS without influencing the security highlights. For taking care of crisis circumstances, crisis office is executed with outer security. Healthcare record management system (HRMS) has created as a tolerant-driven model of well-being data trade. Commonly displaying HRMS organizations for each individual, there are various security and safety risks that may block its huge selection. The focal concern is about whether or not the patients could really control the sharing of their sensitive prosperity realities, particularly when they are made sure about on a public worker which people may not altogether trust. From one perspective, regardless of the way that there exist human administrations rules, for instance, HIPAA which is starting late changed to join colleagues cloud suppliers are commonly not verified substances. Obviously, because of the high assessment of the tricky individual health information, the outcast amassing workers are regularly the objectives of different toxic practices which may impel introduction of individual health information. As an esteemed occasion, a Department of Veterans Affairs information base containing sensitive individual prosperity information of 26.5 million

military veterans, including their administration oversaw investment funds nos. what is more, clinical issues was looted by an agent who took the data outside without endorsement. To guarantee lenient-driven security control over their own HRMS, it is essential to have fine-grained information get the chance to control parts that work with semi-trusted in workers. Therefore, another framework is recommended that guarantees the security of HRMS. That should be done using ABE algorithm. HRMS organization empowers a patient to oversee and control her own one of a kind thriving information in a solitary spot through the web, which has made the breaking point, recovery, and sharing of the remedial data progressively suitable. The principle objective is to verify individual health records utilizing multi expert ABE. That is various experts can get to the HRMS without influencing the security highlights. For taking care of crisis circumstances, crisis office is executed with outer security.

## 2 Literature Survey

### *2.1 Attribute-Based Key Generation Scheme*

- Green et al. prescribed to re-appropriate the decoding load, so the customer can recover the message with lightweight computation. To support the exactness of the changed code text, the evident redistributed decoding issue is inspected in [11] to give a ground-breaking course to the rightness affirmation. To expand the ABE security for video content, a period area trait-based admittance control (ABAC) plan is proposed by Yang et al. [18] to guarantee the cloud-based video substance sharing, which introduces the time into the code text and keys to recognize time control. To decrease the trust of a single position, the multi-pro ABE plan is investigated in [6–8, 10, 13, 16]. Attribute-based encryption (ABE) is another philosophy that reuses the open key cryptography thoughts. Out in the open key cryptanalysis, a message is encoded using the open key shared by recipient. New character-based cryptography changes the standard thought of open key cryptography. It empowers the open key to be a self-assertive string, e.g., the email address or phone no. of the collector.
- ABE goes well beyond and describes the lifestyle as set of attributes not confined to single nuclear key. There are various forms of ABE as multi-authority ABE (MA-ABE), key-policy ABE (KP-ABE), and cipher-text effective access control (CP-ABE), and the public-key encryption (PKE) can be utilized. Be that as it may, there is high overhead of key administration and requires to encrypt numerous duplicates of a document utilizing various clients keys. To give the secured and flexible game plans, one-to-various encryption strategies like ABE can be used. In [12, 17], information is scrambled utilizing credits set with the objective that various customers can decode it. This makes encryption and key administration more effective. In 2016, a secret phrase-based break-glass access plot is proposed,

which is based on two-factor encryption: mystery secret phrase-based encryption and master secret key-based encryption.

### *2.2 Access Control Mechanism*

- The feature-based encryption for data privacy and access control framework proposed in [15] is considered to give data openness in emergency conditions and talks about circulating made sure about information among different client, proprietor, and various position situations. Access control and security of the well-being data in the constant application are given through attribute-based encryption. In [19], maker gives a merged procedure of fine-grained get the chance to order over cloud-based multi-worker data close by a provably secure compact customer affirmation segment which can be endorsed comprehensively in different heterogeneous condition. In [20], author supports facilitated exertion and gathering-based thought movement and ability to utilize applications subject to strategy essentials and an average plan of clinical information. It grants fuse of new organizations reliant on a total and longitudinal point of view on patients paying little mind to where or by whom the thought was passed on.
- Qihua Wang et al. proposed the distributed access control with outsourced computation in fog computing model which relies upon attribute-based encryption (ABE) which can viably accomplish the fine-grained admittance control. In any case, the computational intricacy of the encryption and decoding is developing straightly with the expansion of the quantity of traits [2]. To lessen the computational expense and assurance the secrecy of information, appropriated admittance control without sourced calculation in haze figuring. The standard thought of their methodology is to give fine-grained getting the opportunity to control [14] over sensor information, and it is versatile against ambushes, for example, client enamoring and center settling. Nevertheless, their model relies upon united methodology because solitary the [9, 21] framework regulator can perform key the board.

## 3 Motivation

Patients information is the opportunity not to be watched. It is viewed as a fundamental right that is critical to well-being, security, and personal satisfaction. Maintaining its security and privacy with access control policies for different users is a basic necessity w.r.t to existing applications [1, 3, 18]. However, in current scenario, there is a lack of access control policies as the data can be accessible to only doctors or in other cases the data is made public. In both cases, there are issues of security, privacy, and accessibility gaining all at once. A patient with obscure manifestations could be analyzed and treated with various restorative records in a few emergency clinics in the current medicinal framework. Therefore, recognizing cross-area secure

information sharing framework is important to encourage the treatment of the patient between different medical clinics. For example, medical clinic A's examination report can be obtained by the emergency clinic B specialists. The encrypted therapeutic papers produced by different emergency clinics are sent for capacity and inevitable information to the open mists. Crisis circumstances can occur in restorative frameworks, such as an auto-collision or a sudden swoon to the patient. In these crisis situations, the license's electronic medicinal records are urged to spare their lives. In any case, the rescue vehicle regularly works on the scene with no consent to the encrypted therapeutic documents. In this circumstance, the information protection security tool may prevent the crisis from saving the life of the patient. It is therefore essential for the rescue vehicle faculty to structure a break-glass access method to access the electronic medicinal documents despite the fact that they do not have the associated characteristic mystery key. Meanwhile, to avoid malignant information access by the aggressor, the break-glass access method should be sensible and responsible.

# 4 Proposed System

## *4.1 System Framework*

The proposed model offers a privacy-friendly framework as shown in Fig. 2 data storage and self-adaptive access control system with smart deduplication:

1. **Smart cross-domain data sharing**: Patients may have a spot with various clinical establishments, (for example, one of a kind medical clinics and centers) in certified application. The framework is divided into a few medicinal areas as indicated by medical institutes and backups that medical institutes can safely share a patient's encrypted health documents with social insurance surgeons or specialists from medical institutes. Therapeutic records for patients are encoded using key generation-based attributes and AES Rijndael encryption with a cross-domain access policy, so the approved customers tend to have access throughout the framework.
2. **Smart self-adaptive access control**: In this system, the entrance control segment is self-adaptable with common place and crisis circumstances. Patients and medication staff register to their own clinical area and get the mystery keys quality that can be utilized in run of the mill conditions to acquire scrambled archives for patients. In emergency conditions, a break-glass access methodology is arranged with the target that a break-glass key dependent on a mystery expression can recoup every one of the patient's recorded clinical documents.
3. **Smart deduplication**: This system supports astute deduplication over quality-based scrambled information to save the additional room and diminish the expense of trade between the three-phase public cloud and data customers. It can be tested directed off the bat if the substance of the figure is a real one.

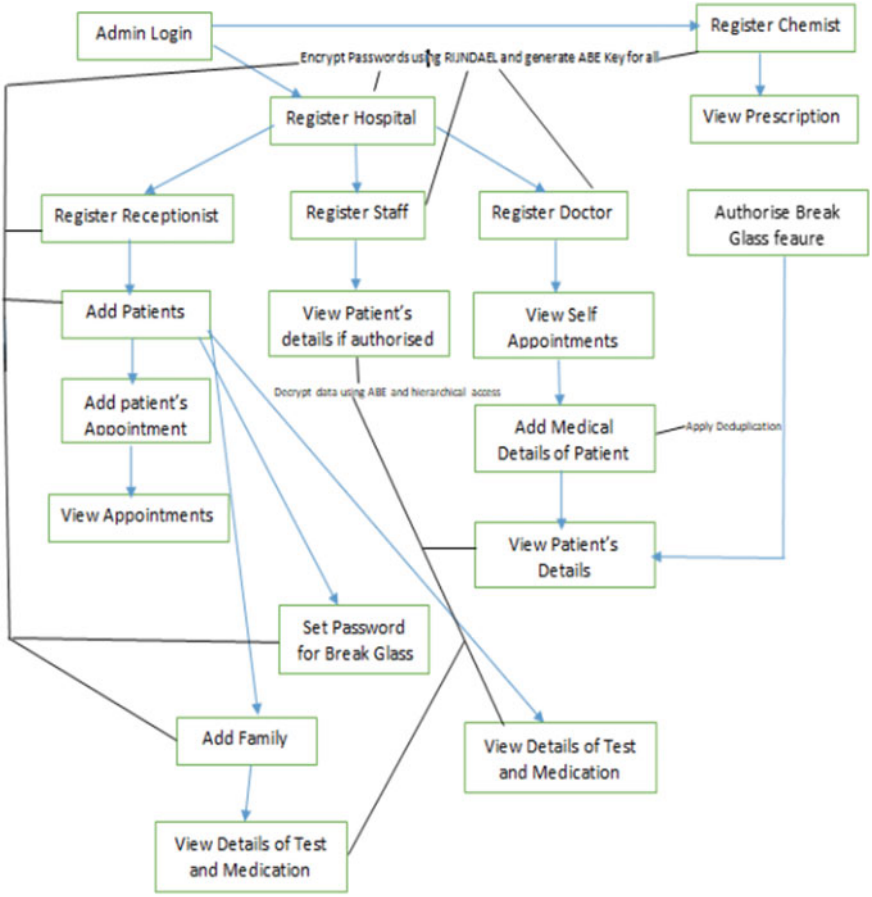



**Fig. 2** System flow

Next, it can be tested whether there is an indistinguishable clinical record in the figure works. In conclusion, the figure content is re-encoded using a consolidated access arrangement with the ultimate goal of getting all the first characterized approved customers from the first figure writings to the encrypted data. No plaintext message is spilled into the public clouds during the deduplication procedure.

## *4.2 System Design*

**Registration** admin has the right to register any hospital and chemist shops to the system. As admin registers the hospitals and chemist, their passwords are encrypted

while storing in cloud, so that the security and integrity of the hospital admin and chemist are maintained. Likewise, hospital admin has the right to register its staff and doctors to HRMS. Their passwords are also encrypted while storing in cloud, so that the security and integrity of the user's password are maintained. Attribute-based key generation is used to generate key of every user for the encrypted data access control. It converts the encrypted document with activation code to the zip file and then sends the activation code to download to the user. For users to access the file, the generated key is important.

**Login** any user who are registered to the application through their respective hierarchy can login through the login page. Password matching is done to the decryption of password and password match.

**Data Upload** This module helps the server to upload a file to the cloud, and the document is encrypted using symmetric encryption algorithm, i.e., AES Rijndael and ABE key generation. After getting logged in the cloud, the doctor can upload a file of patient which he wishes to share. While uploading the file, AES works simultaneously. As the file fragments are finished uploading, the file fragments are encrypted through AES and are spared in encoded design in the cloud. The client who has the position to get to the document would require the trait to coordinate the record access and unscrambling, and exactly at that point, the record can be downloaded.

**Data Access** On the off chance that any client other than the record proprietor, for example persistent, needs to get to the record from the cloud, they have to get to the document with the approved trait key. In the event that the client property key approves the client to get to the document, he can download the record. As the record is spared in the cloud in scrambled arrangement, the document ought to be first brought into its decoded organization to get the record in its unique configuration. As the user downloads the file, the user automatically decrypts the file via the user's attribute key to download the file and the user downloads the file in decrypted format.

**Break-Glass Access** The essential methodology of a hopeful security framework is to accept that any crisis circumstance mentioning information access is authentic and should be conceded. The information subjects are permitted to abrogate explicit access control consents. For this, it has proposed a solid break-glass solution that is dependent on password-based encryption and master secret key-based encryption. We have introduced this system for emergency access of the patient's data by any unauthorized user. This is a break-glass access system based on a password in which the user sets a password for accessing his emergency file. Whenever any emergency situation occurs, anybody can access the patients file who has the password. If the password matches, then all the medical reports of the patient can be viewed. With an enhancement in this break-glass system with respect to its security, the encryption of the password is incorporated. But as the system has a hierarchy-based access control mechanism, it is required to link that encryption with attributes of the person. When the patient sets the password, this password is encrypted and stored in database using AES RIJNDAEL algorithm [22]. In emergency situation, only the hospital admin has the right to decrypt the password using the patients mobile no. The hospital admin

can then provide the emergency access of patients file to the doctors which requires it, and the attribute key is matched of that particular user and doctor. If the key is authorized to break-glass access of the patients file, the doctor can access his file without any access control hierarchy mechanism.

## 5 Performance of Our System

IHRM and the current ABE are contrasted and conspired as far as communication and computation overheads. The tests are coordinated on a PC running Windows 10 64-piece working framework. Figures 3 and 4 show that our system is highly efficient in speed, and as it uses a combination of IBE-ABE, the efficiency of the system with privacy and security is comparatively higher with the other proposed systems for break glass.

IHRM is a giant universe improvement since the expert open key size is predictable and has no limitation on the size of the universe property set. The upside of colossal universe progression is that new credits can be balanced after the structure is set up. Another genuineness is that the extra room is little for assets. The IHRM system is uncommonly capable in speed and moreover can be open and available from any place as it is cloud-based application.

### Algorithm

**AES RIJNDAEL** algorithm uses 10–128-bit keys, which are stored in 4 × 4 tables. The plaintext is also 128 bit chunks stored and divided into 4 × 4 tables. This 128-bit

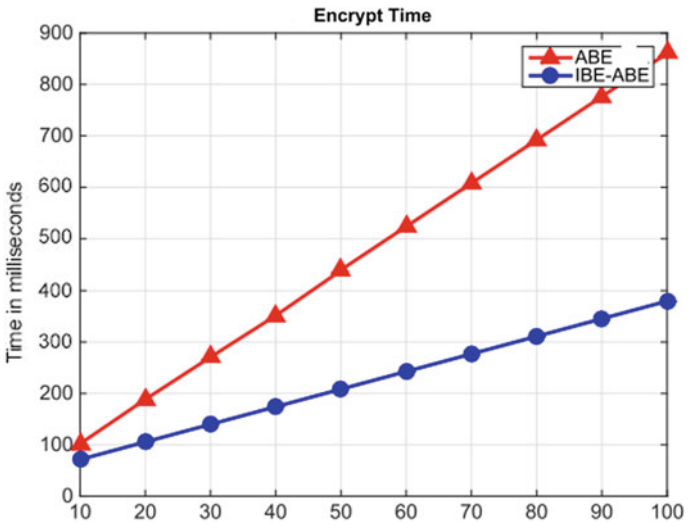



**Fig. 3** Comparsion graph-encryption

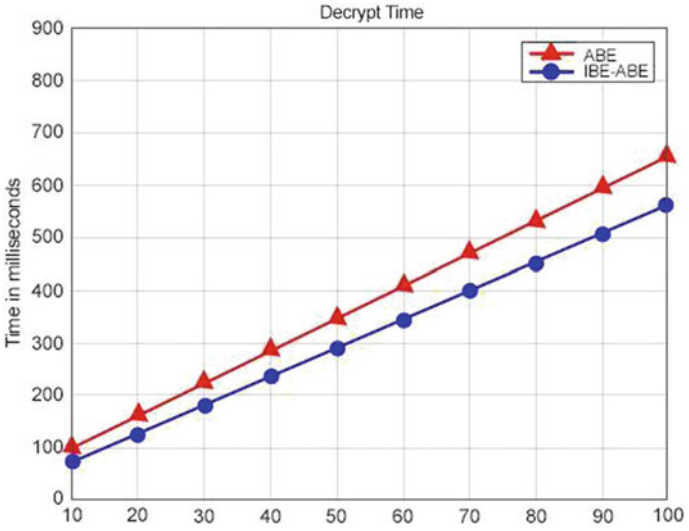



**Fig. 4** Comparsion graph-decryption

plaintext chunk is processed each in a 10-round. The processing rounds depends on the size of the chunk, i.e., 10 rounds on 128-bit keys, 12 rounds for 192 bit, and 14 rounds for 256 bit. The code is generated after the tenth round. Each individual byte is substituted in an S-box and replaced by the reciprocal over Galois field (GF) (2 8). Further, a bitwise application of mod2 matrix along with an XOR operation is performed. Thus, the matrix rows are sorted cyclically followed by exchanging of columns via a Galois field (GF) (2 8) using matrix multiplication technique. In each round, an XOR link is applied to the subkey.

ABE-based key generation: In the key generation algorithm of ABE, there are attributes as follows: x1, x2, … xn0,1 r, rx1, rx2, … rxnRZp are chosen randomly to generate the private key SK where

$$SK = \left[g^{(\alpha+r)/\beta}, g^r \cdot H(x_1)^{r_{x_1}}, g^{r_{x_1}}, \ldots, g^r \cdot H(x_n)^{r_{xn}}, g^{r_{x_n}}\right]$$

Those random rs randomly chosen by the key authority is generated once for all users, and each user will be given the private key based on their attributes. To be more clear, assume user A claims attributes x1, x2 while user B claims x1, x3. They both at different times request a private key containing a component for attribute x1. So, rX1 is the same for both requests as grH(x1)rx1, and grx1 will be affected if they are different.

## 6 Conclusion

The proposed research work has a dispersed stockpiling structure for cross-area remedial archive with break-glass access control and secure deduplication to help the board's data in the medical care framework. Our work focus is on secure deduplication empowering the public cloud to recognize the cipher-texts with a similar therapeutic message; what's more, the restorative foundation's private fog to re-encode the cipher-text with a merged admittance strategy, so all previously approved information clients can get to the new code text. Moreover, this structure is the first to give the clinical applications two suits of access control systems: cross-territory trait-based access in commonplace conditions and break-glass access in crisis circumstances dependent on mystery phrases. These two instruments not only guarantee the safety of scrambled therapeutic documents of patients, but they also provide prompt access to emergency data to the lives of patients.

## References

1. Wu J, Dong M, Ota K, Li J, Guan Z (2018) Big data analysis-based secure cluster management for optimized control plane in software-defined networks. IEEE Trans Netw Serv Manag 15:27–38
2. Wu J, Dong M, Ota K, Li J, Guan Z (2018) FCSS: fog computing based content-aware filtering for security services in information centric social networks. IEEE Trans Emerg Top, Comput
3. Guo L, Dong M, Ota K, Li Q, Ye T, Wu J, Li J (2017) A secure mechanism for big data collection in large scale internet of vehicle. IEEE Internet Things J 4:601–610
4. Wu J, Ota K, Dong M, Li C (2016) A hierarchical security framework for defending against sophisticated attacks on wireless sensor networks in smart cities. IEEE Access 4:416–424
5. Dwivedi AD, Morawiecki P, Wójtowicz S (2018) Finding differential paths in ARX ciphers through nested monte-carlo search. Int J Electron Telecommun 64:147–150
6. Dwivedi AD, Morawiecki P, Singh R, Dhar S (2018) Differential-linear and related key cryptanalysis of round-reduced scream. Inf Process Lett 136:5–8
7. Dwivedi AD, Srivastava G (2018) Differential cryptanalysis of round-reduced LEA. IEEE Access
8. Dwivedi AD, Morawiecki P, Wójtowicz S (2017) Differential and rotational cryptanalysis of round-reduced MORUS. In: Proceedings of the 14th international joint conference on e-business and telecommunications, Madrid, Spain, 24–26 July 2017; pp 275–284
9. Dwivedi AD, Morawiecki P (2019) Differential cryptanalysis in ARX Ciphers, application to SPECK. Cryptology ePrint Archive: Report 2018/899. 2018. Available online: https://eprint.iacr.org/2018/899. Accessed on 9 Jan 2019
10. Dwivedi AD, Morawiecki P, Wójtowicz S (2017) Differential-linear and Impossible Differential Cryptanalysis of Round-reduced Scream. In Proceedings of the 14th International Joint Conference on e-Business and Telecommunications, Madrid, Spain, 24–26 July 2017; pp. 501–506
11. Luo L, Guojun QW (2016) Hierarchical multi-authority and attribute-based encryption friend discovery scheme in mobile social networks. IEEE Commun Lett 20(9):1772–1775
12. Mao et al (2016) Generic and efficient constructions of attribute-based encryption with verifiable outsourced decryption. IEEE Trans Dependable Secure Comput 13.5:533–546
13. Dwivedi AD, Kloucek M, Morawiecki P, Nikolic I, Pieprzyk J, Wójtowicz S (2017) SAT-based cryptanalysis of authenticated ciphers from the CAESAR competition. In: Proceedings of

the 14th international joint conference on e-business and telecommunications, Madrid, Spain, 24–26 July 2017, pp 237–246
14. Maw HA et al (2016) BTG-AC: Break-the-glass access control model for medical data in wireless sensor networks. IEEE J Biomed Health Inf 20(3):763–774
15. Li M, Yu S, Ren K, Lou W (2010) Securing personal health records in cloud computing: patient-centric and fine-grained data access control in multi-owner settings. In: Jajodia S, Zhou J, (eds) Security and privacy in communication networks. Springer Berlin Heidelberg
16. Berlin/Heidelberg, Germany, 2010; pp 89–106
17. Mandl KD, Markwell D, MacDonald R, Szolovits P, Kohane IS (2001) Public standards and patients' control: how to keep electronic medical records accessible but private. BMJ 322:283–287
18. Wu J, Dong M, Ota K, Li J, Guan Z (2018) Big data analysis-based secure cluster management for optimized control plane in software-defined networks. IEEE Trans Netw Serv Manag 15:27–38
19. Sun W et al (2016) Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud. IEEE Trans Parallel Distrib Syst 27(4):1187–1198
20. Yang K et al (2016) Time-domain attribute-based access control for cloud-based video content sharing: a cryptographic approach. IEEE Trans Multimedia 18(5):940–950
21. Shakya S (2019) An efficient security framework for data migration in a cloud computing environment. J Artif Intell 1(01):45–53
22. Suma V (2020) A novel information retrieval system for distributed cloud using hybrid deep fuzzy hashing algorithm. J Inf Technol 2(03):151–160
23. https://lucidoutsourcing.com/blog/cloud-computing

# Malignant Web Sites Recognition Utilizing Distinctive Machine Learning Techniques


**Laki Sahu, Sanjukta Mohanty, Sunil K. Mohapatra, and Arup A. Acharya**

**Abstract** With the development of Web technology, the Internet has become a stage for wide scope of criminal scheme including spam promotion, budgetary fraud, Web page defacement, etc. Because of the fast development of the Internet, Web sites have become the interloper's fundamental objective. As the quantity of Web pages expands, the malicious Web pages are moreover extending and the attack is dynamically gotten progressed. The existing approaches like blacking listing and dynamic analysis approaches are time and resources intensive, hence these methodologies are not adequate to classify the Web sites as malignant or benign. The proposed research work develops a lightweight classification framework to recognize malicious Web pages effectively using distinctive supervised machine learning classifiers like Naïve Bayes, K-nearest neighbors, random forest, AdaBoost, and some distinct relevant URL features. So here, some URL-based features are extracted for the detection of malicious and benign Web sites. We have experimented the test results using Python environment and found the random forest achieves highest classification accuracy of 98.7%.

**Keywords** URL · Classification · Malignant · Benign Web pages · Machine learning classifiers

## 1 Introduction

There are several Web sites that scan any URL for free. Recognition of malicious Web site is one of the most interesting research issues in the field of security. Nowadays, Internet has become a necessary aspect of one's life. The function of Web page's vital

L. Sahu · S. K. Mohapatra
College of Engineering and Technology, Bhubaneswar 751033, India

S. Mohanty (✉) · A. A. Acharya
School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar, India

A. A. Acharya
e-mail: aacharyafcs@kiit.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_39

framework in different area, for instance, ease correspondence, safe fund transfer, proper trade management, and smooth E-administration make education to reach all [1]. One approach to robotization is for Web designers to make sure about and enhance their Web pages. Previous research studies have a few resemblance to this paper and uses different parts of URL features [2]. Blacklisting administration was installed in the programs to confront the difficulties however it has a few detriments like mistaken posting [3]. Malicious URLs are traded-off URLs that are utilized for cyberattacks. To ignore data waste, the URL identification proof is the best arrangement. So the recognition of malicious URL is externally a hot sector of data security because it lunches different types of attacks including drive-by download, social engineering, and spam. By utilizing susceptibilities in modules or include malicious code through JavaScript, drive-by download criticize is typically completed [4]. This paper has presented a lightweight method where the Web pages are dependent on a limited number of features. The proposed research has used four different machine learning (ML) techniques and some relevant features in detecting malevolent Web sites.

The rest of the paper is arranged as: Sect. 2 presents the related work, Sect. 3 describes methodology, Sect. 4 presents experiment and results, and conclusion and future aspects of is described in Sect. 5.

## 2 Related Work

Some correlated machine learning approaches are surveyed for binary classification of malicious Web sites. Tao et al. [5] presented a novel methodology for identifying Web pages as malignant or safe Web pages using classification algorithm (NB, decision tree, and SVM) utilizing HTTP period data (HTTP periods headers, areas of solicitation and reactions) with 50,000 benevolent Web pages and 500 malevolent Web pages. The accuracy of classification algorithms along with trained dataset was 92.2% and with a very low false positive rate was 0.1%. Yoo et al. [6] proposed a two-stage of detection method, in first stage, the discovery model used the decision tree algorithm, for identifying malignant Web pages, in second phase, the recognition model with one class SVM is utilized to identify new kinds of malignant Web pages. The detection rate of the malignant Web pages is 98.9% but the FPR was 30.5%. Dharmaraj et al. [7] presented a features selection and data preparation method in detecting malignant Web pages, detected malicious Web pages statically using URL strings. They used various supervised machine learning algorithms for evaluating the dataset and found the accuracy was between 95 and 99% and very low FPR and FNR for all the models of classification process. Aldwairi et al. [8] designed a lightweight framework considering URL lexical feature, host-based feature, and some special features to recognize malignant Web pages-based. MALURLS framework decreases the training time utilizing genetic algorithm (GA) to extend the training set and get to learn the NB classifiers. TF–IDF improved exactness by up to 10%, JS–Enable–Disable improvement was about 6% while 3–4–5 gm enhancement was restricted to 3%. Wang et al. [9] presented an approach which combines

both static and dynamic analysis for recognizing the malignant Web pages. In static analysis, the static features of pages are identified and the estimators are made to forecast whether the Web page is malignant or safe. But in dynamic analysis context, the Web pages run dynamically in the programming environment and the behavior of the Web pages was analyzed. Kazemian et al. [2] proposed three supervised learning and two unsupervised learning algorithms. They have used all these machine learning techniques to make predictive models for examining more number of malicious and non-malicious Web pages. These Web pages are parsed and different features similarly content, URL, and screenshot of Web pages are machine learning models. A replica of results on supervised learning has built an accuracy of up to 98% and 0.96 for the unsupervised learning technique. Vanhoenshoven et al. [10] presented a malicious URLs detection technique with some machine learning classifiers, such as NB, SVM, multilayer—perceptron, decision trees, random forest, and KNN. They implemented a dataset of 2.4 M URLs and 3.2 M features. The results are obtained using random forest and multilayer perceptron achieves higher accuracy. Sirageldin et al. [11] designed a shell for malevolent pages identification using an artificial neural network learning algorithms. They have considered two features of algorithms such as URL lexical features and page content features. The observation results is a high false positive rate (FPR) which delivered by machine learning approaches is decreased. Abbasi et al. [12] proposed a method where health-related data and guidance accessible on the Web are used for feature vector. They prefer adaptive learning algorithm known as recursive trust lsabeling (RTL). It utilizes substance and chart-based classifiers for increasing recognition of fake clinical sites. The experimental results have shown better accuracy of 94%. The results were obtained for confidence, security, and public safety. Kim et al. [13] build up a Web-Mon instrument which runs quicker than the conventional tools. They have used the techniques RFC, NB, LR, Bayes Net, J48, and SVM algorithms. For a malevolent pages identification method, an active representation was initiated which comprises of WebKit-2, ML, and YARA-based structure.

While there are a few methodologies have been proposed for identifying threatening destinations, the drawback of these approaches is to accomplish their results they used tens and hundreds and thousands of examples, used no process to identify vindictive URL redirection which going up against the difficulties in social event a wide scope of tests. The summary of the literature reviews is represented in Table 1.

## 3 Methodology

This section presents a lightweight binary classification method to recognize the malignant Web pages. The proposed research work has developed a framework of considering different binary classification algorithms including URL-based and host-based features to recognize the malignant Web pages represented in Fig. 1. The dataset is gathered from Web sources for the detection of malicious Web sites. Then, the data is cleaned by eliminating the noisy and messy data. Out of 21 features, only 9 relevant

**Table 1** Summary of literature review

| S. no. | Articles | Techniques used | Feature selection | Results |
|---|---|---|---|---|
| 1 | Tao et al. [5] | Naive Bayes (NB), decision trees, SVM | HTTP session information, domain-based and session header | Detection rate: 92.2% FPR: 0.1% |
| 2 | Yoo et al. [6] | Static analysis | HTML document, JavaScript feature | Detection rate: 98.9% FPR: 30.5% |
| 3 | Patil et al. [7] | SVM, AdaBoost, J48, random forest, random tree, NB, LR, SGD, and Bayes Net | Static features | Detection rate: 95–99% |
| 4 | Aldwairi et al. [8] | Genetic algorithm (GA), NB | Lexical, host-based and special features | Avg. system detection rate: 87% |
| 5 | Woiig et al. [9] | Static analysis, dynamic analysis, hybrid analysis | URL. HTML document JavaScript features | Precision:0_95 Recall: 0.82 F1 score: 0.91 |
| 6 | Kazemian et al. [2] | NB, KNN, SVM, K-means, affinity propagation | Content, URL, javascript features | Accuracy of supervised; 98% unsupervised: 0.96 |
| 7 | Vanhoensho et al. [10] | NB. SVM. Multilayer perceptron. Decision tree. Random forest. KNN. | URL-based features. | Highest accuracy: RFC And Multilayer Perceptron |
| 8 | Sirageldn et al. [11] | Artificial neural network | URL lexical, page content | High false positive rate |
| 9 | Abbasi et al. [12] | Recursive trust labeling (RTL) | Content-based, graph-based features | Accuracy: 94% |
| 10 | Kim et al. [13] | Random forest, logistic regression, Bayes Net, J48, SVM | HTML document, JavaScript and URL features | Detection rate: 98% |

features and 830 records are extracted. Then the dataset is divided into two groups: training dataset consisting of 664 data and testing dataset comprises of 166 data. In step 4: data have been implemented on different machine learning classifiers with the solution of training set. In last step, the machine learning classifiers are utilized to predict the accuracy.

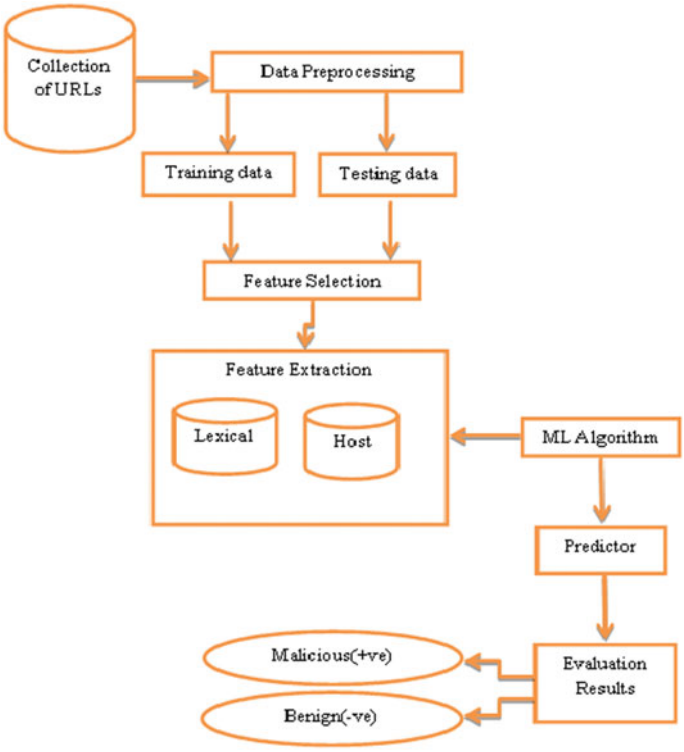


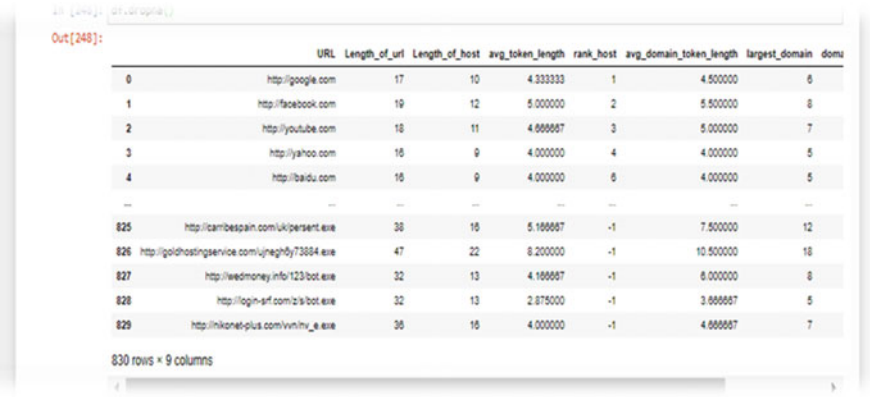

**Fig. 1** Classification framework for malicious Web sites detection

### *3.1 Dataset*

The downloaded Web pages are partitioned into one set yet it contains both malicious and benign [2]. The source for the URL-based dataset accumulated from finding malicious URL through URL features [14]. This dataset has 830 data and 21 features. Out of 830 data, for training set, 664 data are utilized and another testing set 166 data are used. Portrait of our data file is given in Fig. 2.

### *3.2 Selection of Features*

The URL features are selected manually. Out of 21 features, we have selected nine essential relevant features which are required for our research purposes. In the feature selection process, the raw and noisy data are eliminated. These features' description is presented in Table 2. Features are categorized into two groups, i.e., lexical-based feature and host-based feature. Lexical-based feature involves URL, domain token



Out[248]:

| | URL | Length_of_url | Length_of_host | avg_token_length | rank_host | avg_domain_token_length | largest_domain | domu |
|---|---|---|---|---|---|---|---|---|
| 0 | http://google.com | 17 | 10 | 4.333333 | 1 | 4.500000 | 6 | |
| 1 | http://facebook.com | 19 | 12 | 5.000000 | 2 | 5.500000 | 8 | |
| 2 | http://youtube.com | 18 | 11 | 4.666667 | 3 | 5.000000 | 7 | |
| 3 | http://yahoo.com | 16 | 9 | 4.000000 | 4 | 4.000000 | 5 | |
| 4 | http://baidu.com | 16 | 9 | 4.000000 | 6 | 4.000000 | 5 | |
| ... | ... | ... | ... | ... | ... | ... | ... | |
| 825 | http://carribespain.com/uk/persent.exe | 38 | 16 | 5.166667 | -1 | 7.500000 | 12 | |
| 826 | http://goldhostingservice.com/ujneghfy73884.exe | 47 | 22 | 8.200000 | -1 | 10.500000 | 18 | |
| 827 | http://wedmoney.info/123/bot.exe | 32 | 13 | 4.166667 | -1 | 6.000000 | 8 | |
| 828 | http://login-srf.com/zls/bot.exe | 32 | 13 | 2.875000 | -1 | 3.666667 | 5 | |
| 829 | http://nikonet-plus.com/vvn/nv_e.exe | 36 | 16 | 4.000000 | -1 | 4.666667 | 7 | |

830 rows × 9 columns

**Fig. 2** URL dataset

**Table 2** Lexical- and host-based features

| S. no. | Features | Descriptions |
|---|---|---|
| 1 | URL | It is tile anonymous distinguishing proof of tile URL brake down ill tile investigation |
| 2 | LENGTTH-OF-URL | It is a quantity of characters in the URL |
| 3 | Length of host | It needs to be specified as variable in the send |
| 4 | Avg. token length | Number of characters in tile token. The average token length of each approach, the data are taken from the training data |
| 5 | Rank host | It offers services like Web hosting, domain registration, and Web design, Web site development, software development, Web site maintenance on Linux and windows server |
| 6 | Avg. domain token length | Represent the maximum no. of characters you can have in a Web sites address left of "." is 63 characters |
| 7 | Largest domain | That means consider only values of "real numbers" which make tile function defined |
| 8 | Domain token count | It is a simple module to provide alternative domain associated tokens for administers to send e-mails to users |
| 9 | Malicious | It represents target value |

length, largest domain, domain token count, and the rest of the features involved host-based features [14].

### *3.3 Classification Algorithms*

The binary classification algorithms of machine learning are used in our experiments to predict the classification accuracy of malicious Web pages. These are K-nearest neighbor (KNN), Naïve Bayes (NB), random forest (RFC), and AdaBoost classifier.

KNN: KNN is a nonparametric method. It manages marked focuses to prepare name other focuses. It marks the new point by choosing the focuses that are near that point. K is the quantity of nearest neighbors used for accuracy expectation.

Naive Bayes: NB is a binary classification techniques based on Bayes' theorem. A Naive Bayes classifier assumes that the presence of a particular feature in a class is not related to the presence of any other feature.

Random Forest: It commands overfitting utilizing bootstrap aggregator or bagging. The objective of random forest classifiers is to combine various decision trees to get the final result.

AdaBoost: AdaBoost classifier is an ensemble machine learning method used to boost the performance of decision trees on binary classification problems. AdaBoost can be used in conjunction with many types of learning algorithms to improve performance.

## 4 Experiment and Result

The machine learning algorithms such as Naïve Bayes, KNN, random forest, and AdaBoost classifiers are implemented on Jupyter NoteBook [15] and belong to an interactive Python environment. With its integrated support for Pandas, Scikit-learn, Matplotlib and plots a much more understandable presentation of the flow of the code can be designed. Four algorithms are compared to recognize malignant Web pages. But as compared to other algorithms random forest classifier achieves highest accuracy to detect the malicious Web pages.

The different machine learning performance metrics like accuracy, precision, recall, F1 score, and AUC score are used for the detection of malicious Web pages, which is represented in Table 3.

**Table 3** Comparison table of machine learning classifiers

| Machine learning classifiers | Accuracy (%) | AUC score | Precision | Recall | F1 measure |
|---|---|---|---|---|---|
| NB | 96.3 | 0.971 | 0.92 | 1.00 | 0.86 |
| KNN | 93.3 | 0.937 | 0.82 | 0.95 | 0.88 |
| RFC | 98.7 | 0.977 | 1.00 | 0.82 | 0.92 |
| AdaBoost | 96.0 | 0.967 | 0.88 | 1.00 | 0.94 |

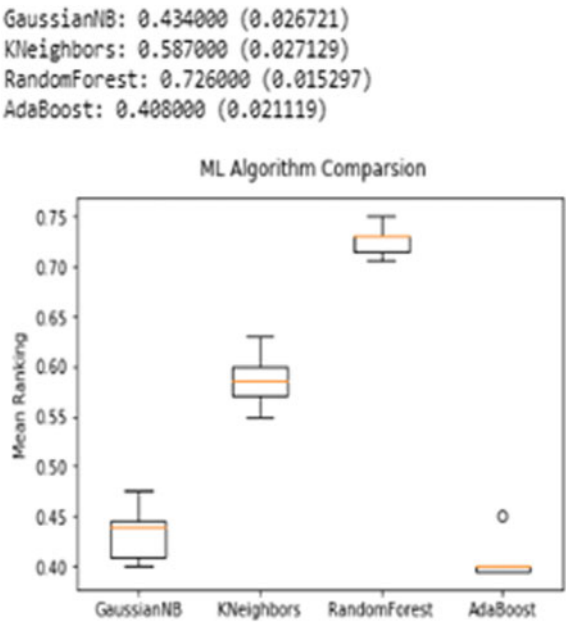



**Fig. 3** Mean and standard deviation accuracy of ML classifier

The mean accuracy and the standard deviation accuracy of different classifiers are depicted in Fig. 3.

The graph in Fig. 4 represents that random forest has the highest accuracy score and gives exact outcome when compared with the other algorithms. It is clearly represented that the RFC classifier achieves highest accuracy of 98.7% and KNN achieves lowest accuracy of 93.3%.

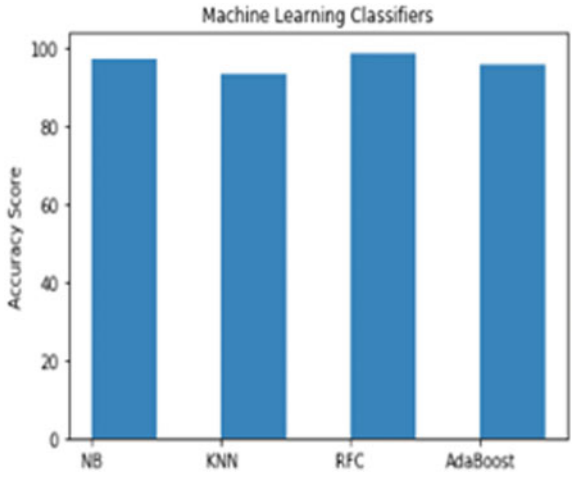



**Fig. 4** Accuracy comparison of ML algorithms

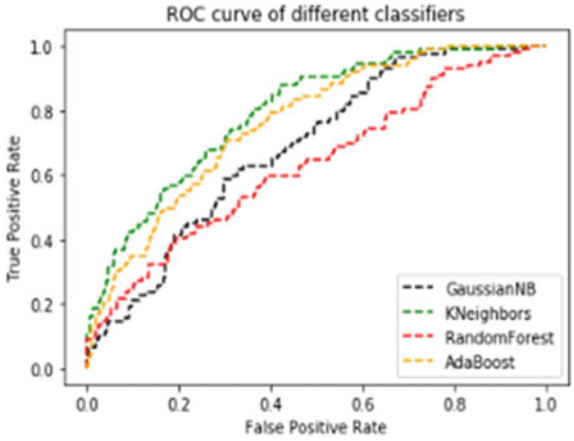



**Fig. 5** ROC curve of different classifier

Receiver operating characteristic (ROC) curve in Fig. 5 is utilized to examine the efficiencies and performance of these four algorithms.

## 5 Conclusion and Future Aspects

The multitude of cyberthreats out there, malicious Web sites play a critical role in today's attacks and scams. Web services are the foundation of connection and communication between applications over the Internet. Recognition of malicious URL is most interesting topic in cybersecurity. Though a few examination contemplates have been performed identifying with the issues of malignant Web pages identification, yet these are expensive as they devor additional time and assets. This paper has utilized new Web site characterization framework dependent on URL-based feature to predict the Web page as malignant or safe utilizing machine learning algorithms. The proposed approach has been implemented and the experimental result shows that random forest classifier (RFC) has the highest classification accuracy of 98.7% of detecting malicious Web pages in comparing various machine learning techniques. In future, it can be planned to use a selection technique for different data sources to improve the performance results of the classifiers.

## References

1. Agrawal J, Agrawal S, Awathe A, Sharma S (2017) Malicious Web pages detection through classification techniques: a survey. IJCST 8(1)
2. Kazemian HB, Ahmed S (2015) Comparisons of machine learning techniques for detecting malicious web pages. Expert Syst Appl 42(3):1166–1177

3. Mohanty S, Acharya AA, Sahu L, Mohapatra S (2020) Hazard Identification and detection using machine learning approach. In: International conference on intelligent computing and control systems. ISBN 978-1-7281-4876-2
4. Lekshmi RA, Thomas S (2019) Detecting malicious URLs using machine learning techniques: a comparative literature review. Int Res J Eng Technol (IRJET) 06(06)
5. Tao W, Shuzheng YR, Bailin X (2010) A novel framework for detect malicious web pages. In: International fourm of information technology and application, vol 2, pp 353–357. IEEE
6. Yoo S, Kim S (2014) Two phase malicious web page detection scheme using misuse and anomaly detection. Int J Reliable Inf Assur 2(1)
7. Dharmaraj R, Patil J (2016) Malicious web pages detection using static analysis of URLs. Int J Inf Secur Cybercrime 5(2):31–50
8. Aldwairi M, Alsalman R (2012) Malurls: a lightweight malicious website classification based on URL features. J Emerg Technol Web Intell 4(2)
9. Wang R, Zhu Y, Tan J, Zhou B (2017) Detecting malicious web pages based on hybrid analysis. J Inf Secur Appl 35(2017):68–74
10. Vanhoenshoven F, Napoles G, Falcon R, Vanhoof K, Koppen M (2016) Detecting malicious URLs using machine learning techniques. In: IEEE symposium series on computational intelligence (SSCI), 1–8
11. Sirageldin A, Baharudin BB, Jung LT (2014) Malicious web pages detection: a machine learning approach. Adv Comput Sci Appl 279:217–224
12. Abbasi A, Zahedi FM, Kaza S (2012) Detecting fake medical web sites using recursive trust labelling. ACM Trans Inf Syst 30(4):22:1–22:36
13. Kim S, Kim J, Nam S, Kim D (2018) WebMon: ML-and YARA-based malicious web pages detection. Comput Netw 137(2018):119–139
14. Ma J, Saul LK, Savage S, Voelker GM (2009) Beyond blacklists: learning to detect malicious websites from suspicious URLs. In: 15th International conference on knowledge discovery and data mining, pp 1245–1254
15. Website: http://jupyter.org/

# Speech Parameter and Deep Learning Based Approach for the Detection of Parkinson's Disease


**Akhila Krishna, Satya prakash Sahu, Rekh Ram Janghel, and Bikesh Kumar Singh**

**Abstract** Parkinson's disease is one of the common chronic and progressive neurodegenerative diseases across the globe. Speech parameters are the most important indicators that can be used to detect the disease at its early stage. In this article, an efficient approach using Convolutional Neural Network (CNN) is used to predict Parkinson's disease by using speech parameters that are extracted from the voice recordings. CNN is the most emerging technology that is used for many computer vision tasks. The performance of the approach is discussed and evaluated with the dataset available in the UCI machine learning repository. The dataset will contain the attributes of voice recordings from 80 individuals out of which 40 individuals are Parkinson affected. Three recordings of each individual are used and 44 features are extracted from each recording of a subject. The experimental setup of the proposed approach on the benchmark data has achieved the best testing accuracy of 87.76% when compared with the available ground truth of the dataset.

**Keywords** Computer-aided system · Parkinson's disease · Convolutional neural network · Deep learning

## 1 Introduction

Researches have been conducted to find appropriate indicators that allow the development of decision support tools that can detect diseases in a very precise and fast manner. Neurodegenerative diseases are caused due to the loss of functionalities of neurons in the brain and peripheral nervous system. Parkinson's disease (PD) is one of the most common neurodegenerative diseases [1]. This disease is characterized by neuron loss in the brain near the motor cortex thus, affecting the movement of a person. As the death of neurons is a very long process, symptoms start to appear at

A. Krishna (✉) · S. Sahu · R. R. Janghel
Department of Information and Technology, NIT Raipur, Raipur, India

B. K. Singh
Department of BioMedical Engineering, NIT Raipur, Raipur, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_40

the later stages [2, 3]. It makes it difficult to diagnose the disease at the early stages and give the proper medication. The most common symptoms of PD are balance loss, muscle stiffness, tremor, and motor skill damage [4].

According to the world health organization, 10 million people are already affected by the disease, and the Parkinson foundation states that by 2020 at least one million people in the US will be affected by PD. As the motor-related symptoms which are the primary measures to diagnose the disease tend to appear only after 80% of neurons die in the brain, the diagnosis of the disease is very harder through motor-related indicators [5]. Besides motor symptoms, one of the most important non-motor symptoms is speech impairment due to the weakening of voice muscles.

The voice disorders are found in 89% of the PD patients [6] and the formation of the abnormalities in the speech is discussed in the publication of James Parkinson [7]. The main abnormalities in speech are due to impaired mimicry, the deficit of laryngeal function, decreased speech force and reduced lung life capacity. These irregularities cause abnormalities in voice including monotonous speech (modulation will be limited), tone of the voice will be lowered, difficulties with changes in loudness, rough and hoarse tone, voltage reduction of vocal folds, change in speech pace, and improper articulation.

A specific clinical technique is not available for the diagnosis of disease instead neurologists predict using medical history through close examination of physical and neurological signs and symptoms of a patient. Computer-aided systems are taking on a new era of disease detection and diagnosis. These technologies ensure a short time diagnosis which can be more efficient as it detects even the slightest variations which are not even perceptible to the human.

The subtle abnormalities in the voice of PD affected people are not perceptible to the listeners at the early stage but can be identified by doing acoustic analyses on speech signals that are recorded [8–10]. The features extracted from the acoustic analysis phase can be used to identify the disorders or abnormalities in the voice. Even though there is no permanent cure to the disease, symptomatic relief can be achieved which in turn helps to enhance the life of patients.

### *1.1 Related Work*

Many kinds of research have been done to prove acoustic analysis can be one of the best non-invasive techniques for earlier detection of Parkinson's disease. The acoustic analysis along with a proper classification framework leads to the identification of the disease accurately. Some of the classification techniques which are employed in this context are artificial neural networks, Support Vector Machines, Regression, etc. [11–13]. Recently deep learning techniques are also found to be effective. In Yasar et al. [11], an artificial neural network was used with voice features extracted from both PD patients and healthy people. Maximum accuracy of 94.93% was recorded in this article using this technique. Hemmerling et al. [12] discusses nonlinear support vector machine to detect the disease using vowel /a/ phonation in a sustained manner and

Principal component analysis (PCA) is used to extract the features. This technique produced a maximum accuracy of 93.43%. The disease detection discussed in Lin et al. [13] is done using a discriminative model based on logistic regression.

Disorders in speech are identified for proper diagnosis of speech-related diseases. Deep learning is found to be the emerging technologies in finding the disorders associated with the voice or speech. The article [14] compared a Multi-layer perceptron algorithm (MLP) and a convolutional neural network architecture for identifying voice pathology or the voice samples with phonetic and speech disorders. Convolutional neural network outperforms MLP with better results. Similarly, another deep learning technique, the Long short term memory model (LSTM) was discussed in [15] for voice pathology detection. The combination of deep learning techniques CNN and LSTM is also an efficient way of voice pathology detection and it is discussed in [16].

This article investigates the effectiveness of the deep learning approach (convolutional neural network) in detecting Parkinson's disease using the vocal features extracted from the voice recordings. Manual selection of the features is avoided and the convolutional layers themselves are used to select the best features that can distinguish between PD patients and normal individuals. Thus, an automatic and easier way of finding the best features is employed. The article is a foundation for further researches on deep learning (convolutional neural network) being an effective framework along with semi-automated and automated feature selections to diagnose the disease with extracted acoustic features.

## 2 Methodology

The proposed deep learning framework is shown in Fig. 1. The first phase of the framework is a data collection phase which is described in Sect. 2.1 of methodology. The next phase is a preprocessing stage where the data is normalized between the range [− 1, 1]. A one-dimensional CNN is built and the data is applied for training the network and the network is tested using another set of data.

### *2.1 Data*

The data used in the experiments of this article was collected from the UCI machine learning repository which contained replicated speech parameters. Three recordings of each subject out of 80 subjects were considered of which 40 are healthy and the other 40 are Parkinson's diseased. 44 voice features or attributes were extracted from each recording which makes the dataset size 44 columns and 240 rows. Voice features from five different families, namely spectral envelope measures, noise features, amplitude local perturbation measures, pitch local perturbation measures, and nonlinear measures are used in the study [8, 9]. Cepstral coefficient of Mel

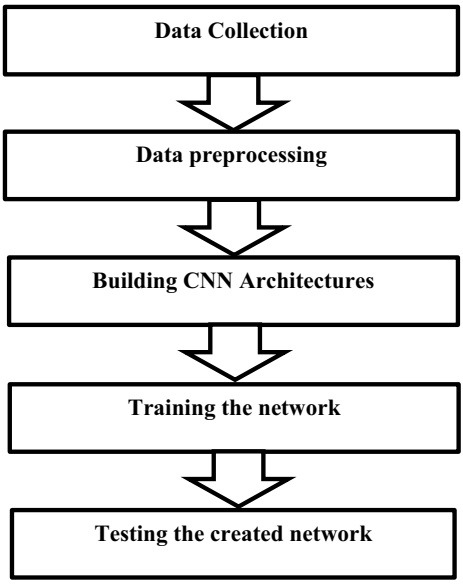



**Fig. 1** Flowchart of deep learning framework

frequency from order 0–12 and their derivatives (MFCC and Delta MFCC) is the spectral envelope measures considered in this study which depends on articular position so the slightest misplacement can also be detected [17]. Harmonic-to-noise ratio (HNR) which measures the ratio of noise present in speech are extracted using voice sauce toolbox and HNR measures are calculated using cepstrum-based technique and Glottal-to-noise ratio (GNE) which measures the ratio of vocal fold's voice excitation oscillation to the excitation by the turbulent noise are the noise features used. Absolute jitter, Relative jitter, average relative perturbation, and quotient of pitch perturbation were considered under the local pitch perturbation family of voice features and it is extracted using waveform matching algorithm. The perturbation measures of amplitude used were shimmer in dB, local shimmer, 5-point amplitude perturbation quotient, 3-point amplitude perturbation quotient, and 11-point amplitude perturbation quotient. The last family was nonlinear measures which have been used in many of the previous researches to find the effect of vocal features in the detection of Parkinson's disease using acoustic features. They include Entropy of Detrended fluctuation analysis, Recurrence period density, and Entropy of Pitch period.

## *2.2 Deep Learning*

It is one of the latest technologies which have got attention from industrial and research area in the last decade. This technology is the key to develop and launch smart products in the market. Multiple layers of processing units are designed using

neural network architecture which uses linear and nonlinear transformation on input data. These can be applied to any kind of data: numerical, text, images, audio, or any combination of them [18]. Deep learning is found to be very effective in identifying the voice disorders in the voice as some of these extracted features can discriminate the people with the disease and the normal one. With the help of deep learning techniques, these features are selected and are used for classification using neural networks [11–13].

**Convolutional Neural Network**

A convolutional neural network or CNN is an important deep learning algorithm that requires comparatively less preprocessing [19]. These are artificial neural networks with alternate convolutional and sampling layers and possess the advantage of self-learning from the given input. The technique is primarily designed for image classification but a variant of CNN called 1D CNN is well suited for one-dimensional data [20]. There are many advantages of 1D CNN over 2D CNN (suited for image classification) which includes low computational complexity due to simple array operations rather than matrix operations, shallow architectures which are easy to design and implement, feasible and relatively fast learning, and is very suitable for low-cost applications [21]. Convolutional Layer, Pooling Layer, and Fully-Connected Layer are the three main important CNN layers.

**Convolutional Layer**

It is an important layer of CNN that performs the feature extraction so that separate manual feature extraction can be avoided reducing the manual interference in the computer automated systems [22]. The layer performs a convolution operation which is very similar to that of the linear operation in a traditional neural network where the input is getting multiplied by the weights which are constantly being updated according to the error generated. In a forward pass, neuron $i$ in layer $l$ receive input from the previous layer as below:

$$\mathrm{In}_i^1 = \sum_{j=1}^{n} \left( W_{ij}^1 x_j + b_i \right) \tag{1}$$

Here $W$ is the weight, $x$ is the input and b is the weight. Accordingly, the two-dimensional array is multiplied with the input data and the two-dimensional array of weights which are called filter or kernel. In specific, the multiplication performed here is dot product and it is done between the much smaller kernel and the kernel-sized part of the input. The output of this operation for a single time will be a single value as multiplication is done element by element and is summed up. The smaller-sized filter will help in multiplying the same kernel multiples times on different points of input. This is done systematically from left to right and top to bottom on the input.

**Pooling Layer**

This is the layer present after the convolution layer and more specifically after nonlinearity is applied on the feature map. It gives an ordering of layers and the layer can

be added multiple times in a model [22]. This layer is applied on each feature map which is produced as an output of the convolutional layer and a completely new different set of pooled feature maps are produced but the number of feature maps will be the same as the previous layer. A pooling operation (like a filter) that has a size lesser than that of the feature map is applied which minimizes the feature map size. The two most common pooling operations are average pooling and maximum pooling. The former will calculate the average of each patch of the feature map and the latter will find the maximum value in each patch of the feature maps.

**Batch Normalization Layer**
The difficulty in taking the right learning rate and initializing network parameters are resolved through this layer. This layer is also used because there is a significant effect by the previous layer parameters on the inputs in each layer which causes a complicated learning problem, covariate shift. Covariate shift is the condition where the inputs of each layer are changing continuously. Feature-wise centering and normalization to mean zero and variance one is the main purpose of the batch normalization layer [23]. Batch normalization has two main stages namely, normalization and scaling. Assuming $f = [f_1, \ldots f_k]$ is the input received, normalization and shift and scale equations are given as

$$\text{Normalization} : f_k' = f_i - \frac{E[f_i]}{\sqrt{\text{Var}[f_i]}} \tag{2}$$

$$\text{Shift and scale} : h_i = \gamma f_k' + \beta_i \tag{3}$$

The new parameters of the network that is used in the training are $\left[\gamma, \beta_i\right]$.

**Fully-Connected Layer** The above-mentioned layers are for finding the suitable features that are capable of classifying the input appropriately [22]. Now the process of classification is done by a Fully-connected layer from the output of convolution and pooling layers. This layer is very similar to the artificial neural network which is used for classification. Before applying the Fully-connected layer, the output from the previous layers is converted into a single vector of values. The probability of the feature vector to be in each class is evaluated and then compared with each other. The most appropriate weights are identified by using their backpropagation process. The weight received by neurons prioritizes the class to which it belongs.

## *2.3 Proposed Architecture of CNN*

This experiment is performed by generating different architectures of the 1D CNN model to identify the disease in a precise way using extracted speech parameters. These architectures are discussed in Table 1. The architectures are built by varying

**Table 1** CNN Architectures used in the experiment

| Architecture no | Architecture 1 | Architecture 2 | Architecture 3 | Architecture 4 (adopted) |
|---|---|---|---|---|
| Layers | Conv1d (8,2)<br>Pooling (2)<br>FC (16)<br>FC (2) | Conv1d (16,2)<br>Conv1d (16,2)<br>Upsampling (3)<br>Pooling (2)<br>FC (16)<br>FC (10)<br>FC (2) | Conv1d (16,2)<br>Conv1d (16,2)<br>Upsampling (3)<br>Pooling (2)<br>Conv1d (16, 2)<br>Conv1d (16,2)<br>Upsampling (3)<br>Pooling (2)<br>FC (16)<br>FC (10)<br>FC (2) | Conv1d (32,3)<br>Conv1d (32, 3)<br>Upsampling (3)<br>Pooling (2)<br>Dropout (0.5)<br>Conv1d (32,3)<br>Conv1d (32,3)<br>Upsampling (3)<br>Pooling (2)<br>Dropout (0.5)<br>FC (16)<br>FC (10)<br>FC (2) |
| Total Parameters | 2762 | 16,912 | 25,136 | 52,498 |

the number of layers and the trainable parameters. As convolutional neural networks do not require a separate manual feature selection, so features are efficiently selected by the model itself. Thus it reduces the effort of handpicking the best features suitable for classification.

The layers of the best performing CNN architecture among all the experimented architectures are shown in Fig. 2. There are two blocks in the architecture where each block is having two layers of convolution, a batch normalization layer, and a dropout layer. The input is passed into the two blocks and the output of the second block is flattened to a single vector of values. This vector is given to two layers of fully connected layers for classification. Training and testing are performed to validate the model developed. The training phase is carried out by dividing the data into two groups training set and testing set where 20% of the total data is divided and kept for testing the model and the rest is applied to train the network model for increasing the classification accuracy.

## 3 Experimental Results and Discussion

The proposed convolutional neural network is applied to the voice feature dataset. Python programming language was used to implement with Keras library on a computer with the configuration of Intel core i3 processor with 4 GB random access memory. The performance of the classifiers is evaluated using measures such as classification accuracy, sensitivity, and specificity. The basic reliability of the classifier is determined by accuracy. The sensitivity and specificity measure how well the classifier predicts one of the two categories.

The training and testing accuracy of the architectures referred to in Table 1 is compared in Table 2. It is visible that architecture 4 produces a maximum accuracy

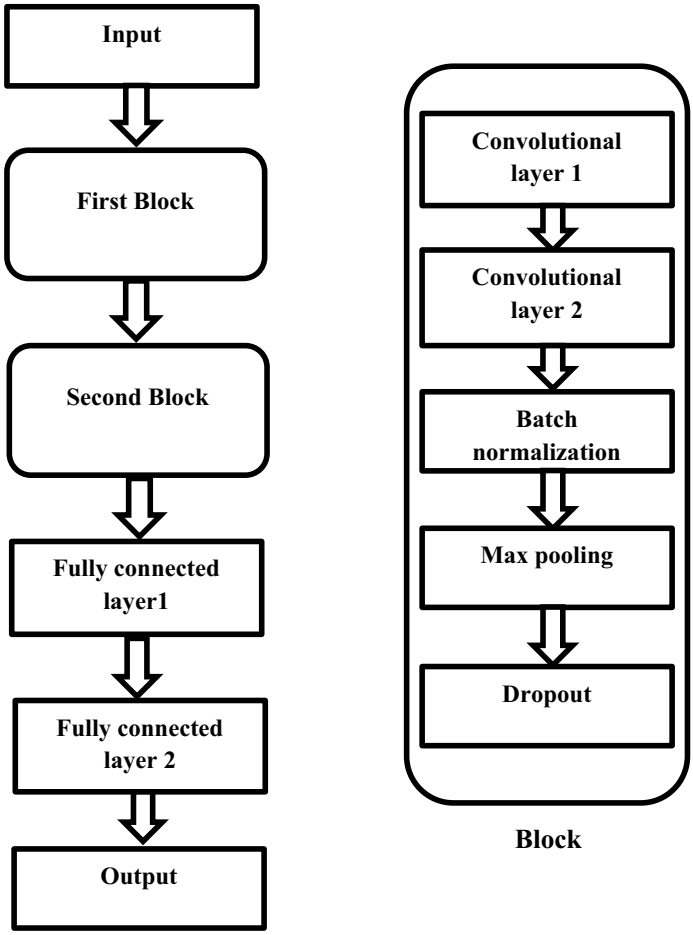



**Fig. 2** Adopted architecture of CNN

**Table 2** Performance of different architectures

| Architecture No | Training accuracy (%) | Testing accuracy (%) |
|---|---|---|
| Architecture 1 | 79.63 | 77.97 |
| Architecture 2 | 86.42 | 82.87 |
| Architecture 3 | 87.28 | 85.69 |
| Architecture 4 | 89.13 | 87.76 |

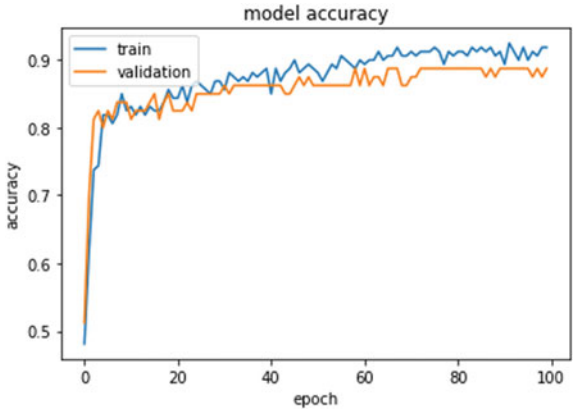



**Fig. 3** Training and testing accuracy of architecture 4

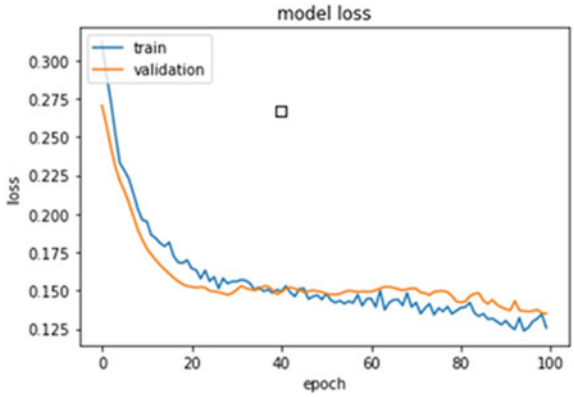



**Fig. 4** Training and testing loss of architecture 4

of 87.76%, a sensitivity of 88.7%, and a specificity of 90.5% when architecture 4 is applied with a small dataset of speech parameters. The performance of the model is depicted in Fig. 3 where it is visible that the training accuracy and testing accuracy are almost the same and don't differ too much. Figure 4 shows the loss of the model and this also states that the training loss and testing loss also don't differ much and the value of loss in the last iterations is reaching 0.125.

## 4 Conclusion

Diagnosis of the disease using a computer-aided system has made revolutions in the medical field. The proposed model is to diagnose Parkinson's disease with a convolutional neural network in diagnosing the disease using speech parameters. Even though the dataset available for the research was having a limited number of entries, a classification accuracy of 87.76%, sensitivity of 88.7% of and specificity of

90.5% is obtained with architecture 4. The results are very promising in comparison to the size of the data used in the research. The proposed architecture is a very simple way of diagnosing the disease with minimal layers of convolution. Compared to any other ways of identifying the disease, voice features enable a low-cost, efficient, and easy-to-use technique and it is very handy as there is no separate feature selection phase. The results are positive to think that on large datasets with more number of subjects, deep learning will outperform the results produced by machine learning algorithms. Hybrid models can be used to enhance the results in the future.

## References

1. Grover S, Bhartia S, Yadav A, Seeja K (2018) Science direct predicting severity of Parkinson's disease using deep learning. Procedia Comput Sci 132(Iccids):1788–1794
2. http://www.pdf.org/speech_problems_pd
3. Lang, AE, Lozano AM (1998) Parkinson's disease. New England J Med 339(16):1130–1143
4. Benba A, Jilbab A, Hammouch A (2016) Analysis of multiple types of voice recordings in cepstral domain using MFCC for discriminating between patients with Parkinson's disease and healthy people. Int J Speech Technol
5. Orozco-Arroyave JR, Hönig F, Arias-Londono JD, Vargas-Bonilla JF, Skodda S, Rusz J, Nöth E (2015) Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinsons disease. In: Anual conference of the speech and communication association (INTERSPEECH), pp 95–99
6. Wodzinski M et al. (2019) Deep learning approach to parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. In: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE
7. Parkinson J (2002) An essay on the shaking palsy. J Neuropsychiatry Clin Neurosci 14(2):223–236
8. Naranjo L, Pérez CJ, Campos-roca Y, Martín J (2016) Addressing voice recording replications for Parkinson's disease detection. 46:286–288
9. Naranjo L, et al (2017) A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. Comput. Methods Programs Biomed. 142:147–156
10. Bocklet T, et al (2011) Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis. In: 2011 IEEE Workshop on Automatic Speech Recognition & Understanding. IEEE
11. Yasar A et al (2019) Classification of Parkinson disease data with artificial neural networks. In: IOP Conference series: materials science and engineering. In: IOP Publishing, vol 675, no 1
12. Hemmerling D, Sztaho D (2019) Parkinson's disease classification based on vowel sound. In: Proceedings of the 11th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications
13. Lin J (2019) A preliminary study on Parkinson's disease with regularized logistic regression method. Open J Soc Sci 7(11):126–132
14. Kanimozhiselvi CS, Balaji Prasath M, Sathiyawathi T. Voice pathology identification using deep neural networks
15. Gupta V (2018) Voice disorder detection using long short term memory (lstm) model. arXiv preprint arXiv:1812.01779
16. Harar P et al (2017) Voice pathology detection using deep learning: a preliminary study. In: 2017 International conference and workshop on bioinspired intelligence (IWOBI). IEEE

17. Upadhya SS, Cheeran AN, Nirmal JH (2018) Biomedical signal processing and control Thomson multitaper MFCC and PLP voice features for early detection of Parkinson disease. Biomed Signal Process Control 46:293–301
18. Pereira CR, et al (2016) Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. In: 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE
19. Faust O, Hagiwara Y, Jen T, Shu O, Acharya UR (2018) Computer methods and programs in biomedicine deep learning for healthcare applications based on physiological signals: a review. Comput Methods Programs Biomed 161:1–13
20. Kalchbrenner N, Grefenstette E, Blunsom P (2014) A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188
21. Frid A, et al (2016) Diagnosis of Parkinson's disease from continuous speech using deep convolutional networks without manual selection of features. In: 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE). IEEE
22. Pereira CR, Pereira DR, Papa JP, Rosa GH (2016) Convolutional neural networks applied for Parkinson's disease identification. 2:377–390
23. Liao Z, Carneiro G (2016) On the importance of normalisation layers in deep learning with piecewise linear activation units. In: 2016 IEEE winter conference on applications of computer vision (WACV). IEEE

# Study on Data Transmission Using Li-Fi in Vehicle to Vehicle Anti-Collision System


**Rosebell Paul, Neenu Sebastian, P. S. Yadukrishnan, and Parvathy Vinod**

**Abstract** This paper examines the relevance of a fast approaching highly secure and fast data transmission technique using Li-Fi. It describes the upcoming technology Li-Fi and its applications as well as the developments made in it so far. It enlightens on the new era that will soon be used in almost all domains like health sector, school, bank and so on. An application framework design has been studied to analyze the role of Li-Fi in the process of communication.

**Keywords** Li-Fi · Light emitting diode (LED) · Wireless communication · IoT · V2V communication · Data transmission

## 1 Introduction

Light emitting diodes (LED) bulbs that are used in our household as light source are not only capable of lighting the surroundings but can also be used to transmit data. This idea forms the backbone of the new technology Li-Fi. Li-Fi stands for light fidelity and it uses LEDs intensity variation for transmission of data. It is very difficult or almost impossible for the human eyes to trace this variation as it happens too rapidly. We can say that Li-Fi is a light-based Wi-Fi technology. The Li-Fi revolution can possibly eliminate most of the problems of the existing wireless-fidelity infrastructure (Wi-Fi) in several core areas like medicine and health sector

R. Paul (✉) · N. Sebastian · P. S. Yadukrishnan · P. Vinod
Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Karukutty, Kochi, India
e-mail: rosebell@scmsgroup.org

N. Sebastian
e-mail: neenusebastian@scmsgroup.org

P. S. Yadukrishnan
e-mail: yedhups@ieee.org

P. Vinod
e-mail: parvathyvinod@ieee.org



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_41

since it evokes no electromagnetic reaction. The recent studies made in Li-Fi [1] reveal that it is going to be a part of the framework of 5G. It has been recognized by IEEE since the end of 2017, as an IEEE 802.11 working group has been formed for making the studies on standards for visual light communications. Wi-Fi has to abide to the country's particular regulations as it works with the spectrum allocation but Li-Fi does not.

Professor Harald Haasa German physicist, known as the founder of Li-Fi, introduced the term Li-Fi and he is the co-founder of PureLi-Fi which is a company based on light communication established by him in 2012 to work intensively on this new finding. He successfully illustrated a Li-Fi prototype at the TED Global conference in Edinburgh on 12 July 2011. He used a table lamp with an LED bulb to transmit a video that was then projected onto a screen. [2]. This idea pioneered by him gave a new direction to the researchers to explore more on the light waves [3].

In the next section, the basic principle of Li-Fi is explained followed by a briefing on a few applications of Li-Fi with several examples where it has already been implemented. In the Sect. 4, a comparative study is made with Wi-Fi. Section 5 describes the market growth of Li-Fi along with a study of converging Li-Fi with several emerging technologies. Section 6 is a prototype model framework illustrating the data transmission using Li-Fi in a vehicle to vehicle communication to avoid collision. Sections 7 and 8 show the analog and digital data transmission in our system. The results and inferences obtained from our experimental study are mentioned in Sect. 8. Finally, this paper is a journey through the developments made so far using Li-Fi and the authenticity of using Li-Fi for data transmission in analog and digital format is studied using the prototype model.

## 2 Principle of Li-Fi

Li-Fi can be termed as light-based Wi-Fi, i.e., instead of radio waves, it uses light to transmit data. In place of Wi-Fi modems, Li-Fi would use transceivers that have LED lamps that could brighten a room as well as transmit and receive information. It makes use of the visible portion of the electromagnetic spectrum which is hardly utilized.

LED bulbs or any light source can be used to transmit the data and the photodetector is used to detect the flash light emitted from the transmitter side. The data to be transmitted has to be encoded into a binary format which consists of a sequence of 1's and 0's. This binary stream of data is given as input to the LED light source and it will be transmitted when the LED glows. These flashes are then detected by the photosensors. The photosensors transfer the binary data for amplification, thereby strengthening the signal in order to decode the binary digits. Finally, the decoded data is transferred to the end device [4].

The block diagram of a Li-Fi transmitter and receiver system is depicted below.

Transmitter section consists of the Web server, modem and LED driver (Fig. 1).

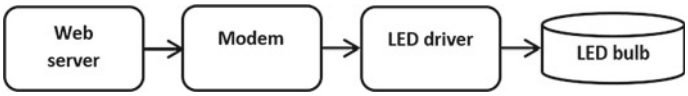



**Fig. 1** Basic block diagram of the transmission section

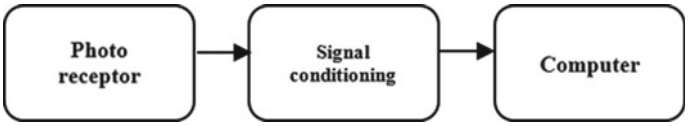



**Fig. 2** Basic block diagram of the receiver section

Modem performs the necessary modulation to the incoming data, thereby making the data signals more suitable. As we change the intensity of the current send to the LED bulb, the brightness of the bulb changes which can then be converted into 1 s when illumination is maximum and 0 s when illumination is minimum. In this way the LED bulb is rapidly flickered, millions of time and is not visible to our naked eyes. In order to control this flickering properly, we make use of modulation. The three different types of modulation in Li-Fi are: single carrier modulation (SCK), multiple carrier modulation (MCK) and color modulation [2].

LED driver changes the driving current for this LED bulb according to the incoming streaming data to get the flickering. It plays the role in the conversion of digital signal to photonic signal.

Receiver section consists of the photoreceiver and is then given for signal conditioning before being transferred to the computer (Fig. 2).

Photoreceiver or the photodiode converts the optical signals received into electrical impulses. Signal conditioning mainly performs the necessary amplification and processing of the received signals before sending to the work devices. Thus, we can say that Li-Fi provides a one to one point of communication between any two devices in a highly secure manner. It has a speed much faster than the existing connection technologies. Though the term came into being in 2011, several studies and research are still being done to make it accepted all over the world. Since it makes use of the visible light spectrum, it can be used to pave way for communication at places where Internet access is denied like petrol pumps, operation theater, aircraft, etc. A study on the several applications developed using Li-Fi in several systems reveals that this technology will soon conquer the whole world. Li-Fi can protect the mankind from radio waves as the light spectrum used in Li-Fi does not penetrate through the body and is much safer when compared to the hazards caused by radio waves [5].

# 3 Applications of Li-Fi

## *3.1 Li-Fi in Medical Department*

The radiation hazards caused to the human body can be drastically reduced if we make use of Li-Fi technology. The minute details of patients can be continuously monitored as long as there is a LED light in the room and the data transmission is done in such fast rate that emergency situation can be handled well. The main challenges in our hospitals are that the rooms require multiple access, and therefore the network becomes dense and heavy and complex cables are needed for the different equipment. The typical network architecture which consist of laying cables, installing routers, switches, etc., becomes difficult in hospitals. Schnell Li-Fi along with Huawei has come with the idea of all optical hospital in which they use Li-Fi and optic fiber as a better choice for hospital infrastructure. They called it green networks as Li-Fi technology uses passive light with no known health hazards. Similarly, Li-Fi is the best option in a room which has confidential information as it cannot be hacked. The photonics network will always be preferred in future for sensitive environment which consist of several medical devices as the Wi-Fi signals might interfere with the device signals at hospitals thereby affecting the readings. The various devices in the hospitals are connected to each other and may also require to transfer data to cloud for future analysis. As IoT enters the surgery room, Wi-Fi will become obsolete. The Li-Fi devices of Schnell were proven to be compatible with the Huawei and so they just had to plug and play making it easy to assemble and reconfigure. Thus, Li-Fi provides wireless connectivity for telemedicine, lifesaving applications, hospital security systems and vital body sensors. Similarly, Li-Fi indoor positioning system can optimize the navigation within the hospitals [6].

## *3.2 Li-Fi in Aircraft*

For the first time in world, Air France A321 tested Li-Fi Technology during a flight in a 12-seater specially designed aircraft in which the travelers could play video games. By working on it for a short time span of a little of 6 months they built the system and successfully used Li-Fi describing it to be a much simpler technology than the traditional equipment they used and hosted the Air France Trackmania Cup in October 2019. They used fiber optic cables to transmit the light-based Internet access. They installed a baggage compartment which acts as the server that was designated to transmit the data notably the video games. They had installed a dedicated headrest for the screen and a dongle for transmitting Li-Fi communications. A detector was installed for optical transmission. Thus, Li-Fi enabled 10 gamers to play together and it is several times faster than Wi-Fi, thereby provided a more viable experience.

The ability to safely access the Internet at 30,000 ft is becoming commonplace nowadays. However, as passengers place greater demands on the in-flight Internet

bandwidth, the connected experience could become less of a benefit to the experience and more of a frustration, because airplane mode blocks your connection to cellular networks and it is not possible to communicate or make calls using phone calls. The airplane mode will turn off the Wi-Fi connection and thus will abrupt all the services that requires Internet connectivity.

Li-Fi solves a 'congestion' issue. In the data-driven world that we live in, we are running out of radio spectrum. This is a problem in crowded places like airports and aircraft interiors as it means that the current available bandwidth does not support the hundreds of people wishing to use data-intensive applications and the Internet in the same place at the same time. Li-Fi solves this issue by using 1000 times the bandwidth compared with the entire radio frequency spectrum. This is additional free, unregulated bandwidth in the visible light spectrum.

Li-Fi paves the way for local area networks to be established, which means that passengers can make calls, use the Internet and access in-flight entertainment systems more easily.

In a world fast being dominated by big data, safeguarding information is paramount. In the cabin, although Li-Fi signals can leak through windows, the technology offers greater protection to passengers than a Wi-Fi connection. But the biggest gains will be experienced by OEMs. Their manufacturing halls often have lots of LED lighting and few windows, which will enhance data security in their facilities.

### *3.3 Underwater Communication System Using Li-Fi*

The underwater communication system is a research area of great demand as the limitations in bandwidth, energy consumption of the devices used and distance pose a big challenge. The underwater communication system based on Li-Fi technology provides protection against ship collisions on the sea. Most remotely operated underwater vehicles (ROVs) are controlled by dedicated wired connections. Since light can travel through water, Li-Fi based communications could offer much greater mobility. An underwater communication system utilizing visible light communication (VLC) technology will be ideal for military and scuba divers operating under vessels, allowing communication within the light spectrum to a certain radius via an audio system built into their diving suits. This project focuses on the safety of the sea in which the headlights, which consist of LEDs acting as transmitters, communicate with photosensors acting as receivers. White LEDs used in the head and tail lights can effectively be used for short-range communication with the photodetectors. The application is cost effective as LEDs are cheap and simple algorithms are proposed for signal generation and transmission.

A study on the usage of Li-Fi for underwater communication by observing the propagation of visible light which is affected by inherent optical properties absorption and scattering is made in. It describes the bio-optical model which concentrates on the chlorophyll approach of propagation of light in seawater [7].

Recently, SaNoor's Laser Li-Fi Solutions used laser-based rays underwater for data transmission and to link several devices and allow Internet of underwater things. They claim that this technology enabled them to allow transmission at giga bytes per second. This system removed all the hassles created in the several existing underwater data transmission systems which consists of lot many of wires as well as and has limited operational range. The signal was sent from seafloor passing through sea and air in spite of the constant motions of the sea waves as well as the foams and bubbles. They found this solution useful for many applications like underwater pipe leak monitoring, underwater environment monitoring, offshore wind fan, etc.

### *3.4 Li-Fi in Office Rooms*

PureLi-Fi came up with a fully networked Li-Fi system by the end of 2017. In the Li-Fi-enabled boardroom, they equipped six lights and it was more than enough for any kind of uninterrupted streaming services. They had come up with a dongle in 2016 which can be connected via universal serial bus to the laptops or other devices on which we need the data transfer and Internet connectivity. This dongle consists of a photodiode and the Infrared transmitter which is used to transfer data back to the LED source. PureLi-Fi had then revealed their future plan of making of semiconductor technology to neutralize this usage of dongle and make Li-Fi core-integrated.

BeamCastor is a Li-Fi-based local area network which was developed by a Russian company Stins Coman in 2014. In this, they made use of a router which was capable of transmitting signal using a light beam that had a coverage of approximately 7–8 m. It was successfully transmitted to eight devices at same time placed at different positions in the office and achieved a four times faster rate than Wi-Fi. The transmission model has to be installed on the ceiling and the receiver model is configured on the work devices. The key highlights of this system were the mobility and the ease of configuration and it could be easily disassembled.

### *3.5 Li-Fi in V2V and V2I Communication*

Vehicle to vehicle communication (V2V) has an underlying wireless protocol called dedicated short-range communications (DSRC). A lost cost V2V communication system is developed DSRC and global positioning system (GPS). The vehicle controlling information such as transmission rate, brake status, speed, acceleration, global positioning and path history are taken into consideration to prevent crashes. With the help of these information if a vehicle in front puts a sudden brake suitable transmission of warning signal can be sent. Similarly, if a vehicle in front makes a slow movement and the vehicle you are driving is moving at pace faster, it might cause a

collision. V2V communication helps to pass the alert on time using the wireless communication. Li-Fi technology can be clubbed suitable in this system as it is having high speed transmission rate which will help in this time critical situations to prevent accidents [8].

In case of traffic system, Li-Fi can be used to enable the vehicle to infrastructure communication (V2I) as the LED in the traffic lights can act as the medium to pass signal to the first car in the lane. This car can then transmit this information to the one behind it. The one to one point communication technology principle of Li-Fi helps to ensure secure communications without congestion. The usage of Li-Fi in smart traffic net which is soon going to be the upcoming transportation system has been described in [9].

## 4 A Comparative Study Between Li-Fi and Wi-Fi

### *4.1 Wi-Fi Technology*

Wi-Fi is an existing wireless networking technology which was invented by NCR corporation in 1991 and is widely in use. By using this technology, we can exchange the information between two or more devices. The radio waves act as the medium thereby making the Wi-Fi networking possible. Wi-Fi uses routers and radio frequency and the radio signals are transmitted from antennas and routers. These signals are picked up by Wi-Fi receivers, such as computers and cell phones that are ready with Wi-Fi cards. Wi-Fi is used for Internet browsing with the help of Wi-Fi kiosks or Wi-Fi hotspots. Wi-Fi data transfer speed ranges from 150 to 2 Gbps. It has a coverage area of up to 32 m Wi-Fi radiations that can harm human health. It is not suitable for airlines and undersea communications.

### *4.2 Comparison with Li-Fi System*

(a) System Requirements: Li-Fi only needs lamp driver, LED bulb and photodetector which are not at all expensive. Wi-Fi requires routers to be installed and subscriber devices are referred as stations.
(b) Capacity: The visible light spectrum is several thousand times larger than the spectrum of radio waves. For visible light, the frequency ranges from 400 to 700 THz, but for radio waves, the range is from 3 kHz to 300 GHz. Thus, Li-Fi signals which are based on visible light spectrum will definitely have a larger capacity.
(c) Efficiency: LED lights consume less energy and are highly efficient. Li-Fi can successfully work in a high dense environment, whereas Wi-Fi can flawlessly work in less dense environment due to interference related issues.

(d) Availability: Light sources are present in every nook and corner of the world. Hence, the availability is not an issue. The billions of light bulbs worldwide need only be replaced by LEDs.

(e) Interruption: Light of course does not cross the walls and thus data transmission using light waves is more secure than Wi-Fi where the radio waves can penetrate across the walls. Thus, it is not easy to intercept the data transmitted via light beam, but hackers can eavesdrop the data transmission using Wi-Fi. On observing the electromagnetic spectrum, the bandwidth of the radio waves is up to 300 Ghz. The bandwidth of the visible spectrum alone is up to 300 Thz which approximately thousand times more compared to the bandwidth of the radio waves. Thus, apart from attaining higher speed, there will not be any electromagnetic interference.

(f) Data Transfer Speed: 4G LT and LTE advance have a maximum of 1 Gbps for stationary users and 100 mbps for mobile users which are theoretical figures and the actual speed might be lesser than these. The average 4G LTE speed provided by countries with highest Internet usage is 40–50 Mbps. Researchers have found up to 10 Gbps in Li-Fi prototype models. This speed can be increased more by using array of LED source or by using the multiple color RGB LED's where each of it can transfer at a speed of 10Gbps. In such a manner up to 100Gbps, when the whole visible spectrum can be used. Wi-Fi standards IEEE 802.11.ac have a maximum 1.3 Gbps transfer speed which comes under 5 GHz band and IEEE 802.11.ad provide a maximum speed of 7Gbps as defined theoretically. Though these speeds are enough to meet our daily need now, as we step into the era of Internet of things, these might not be enough and there comes the urge of maybe 100 Gbps where Li-Fi comes into effect. Thus, Li-Fi is ten times better than the current Wi-Fi on the basis of speed [10].

(g) Coverage: Li-Fi coverage depends on the intensity of the light but expected coverage is 10 m. Wi-Fi provides about 32 m (WLAN 802.11b/11 g) and it will vary based on the transmit power and antenna type.

(h) Technology: Li-Fi present Infrared Data Association (IrDA) compliant devices and Wi-Fi presents WLAN 802.11a/b/g/n/ac/ad standard compliant devices [11].

| Specification | Li-Fi | Wi-Fi |
|---|---|---|
| System requirement | LED bulb, photodetector | Router, subscriber devices |
| Capacity | 400–700 THz | 3 KHz–300 GHz |
| Efficiency | Consume less energy and highly efficient | Affected by inference and less efficient |
| Availability | Light sources and widely available | Need special requirements |
| Data transfer rate | Very high (~1Gbps) | Low (100 Mbps–1 Gbps) |
| Cost and power consumption | Low | High |

(continued)

(continued)

| Specification | Li-Fi | Wi-Fi |
|---|---|---|
| Coverage | ~10 m (depending upon the intensity of light) | ~32 m (depending upon the transmitting power and antennae type) |
| Technology | Infrared data association (IrDA) compliant devices | WLAN standard compliant devices |

## 5 Li-Fi in the Industry and Convergence of Li-Fi with Emerging Technologies

There are many companies that have started to bring the Li-Fi-based products into the markets.

PureLi-Fi, co-founded by Professor Harald Haas, provides kits to support the evaluation of high speed Li-Fi components, light antenna technology for integration into the mobile devices and the systems that allows to deploy Li-Fi.

Oledcomm is providing Li-Fi network interface devices. And they are concentrating on improving the design of Li-Fi router solutions for LED-based systems.

LightBee is working on automotive control access and car2car communications modules.

Velmenni, Firefly, Lucibel, LightBee, LVX System, Signify, Vlncomm, LIFX, Luciom are a few of them which are all concentrating on Li-Fi technology. According to the survey made by marketresearchfuture.com, Li-Fi market is expected to grow at approx. USD 51 Billion by 2023, at 70% of CAGR between 2017 and 2023. Because of the revolution in the technical world where all the devices are getting connected high speed of data transmission is going to be a matter of prime concern and foreseeing this many companies like General Electric and Philips are investing into Li-Fi market.

### *5.1 Li-Fi and IoT*

Internet of things is a system of interrelated computing devices which can exchange information and transmit data without human interaction. IoT devices comprises of several sensors, actuators, connectivity/communication electronics and software to capture, filter and exchange data about themselves, their state and their environment. However, this widespread development in IoT which interconnects larger scale of heterogeneous devices creates a big challenge so as to how to safeguard the hardware and the networks in the IoT system and transfer data without interferences and delay [12]. Thus, integrating the Li-Fi concept with IoT provides solutions to a wider

variety of problems in different domains. Unipolar orthogonal frequency division multiplexing (U-OFDM) which is used in Li-Fi technology gives high speed data transmission along with room illumination. As a result, enough bandwidth is obtained to accommodate large number of IoT devices [13]. Massive multiple input multiple output (MIMO) which consists of several antennas at the transmitter and receiver based on visible light spectrum will have higher bandwidth [14].

### *5.2 Li-Fi and Cloud*

Cloud acts a centralized storage and it allows to store, manage and process data over the Internet using large group of servers based on virtualization technology. The data is stored and retrieved in a data center, instead of locally on the user device. As IoT-enabled environment requires large amount of data to be stored and processed frequently, cloud makes it possible to enable access of the same data and applications from any part of the world. The increasing demand of wireless communication has posed a problem of acquiring real-time data without latency delay and bandwidth bottleneck problems from cloud. In the existing IoT network, especially in highly dense environment, the available radio spectrum becomes insufficient. Li-Fi due to its high speed and greater bandwidth can be a stand in such a scenario. The challenges of the existing cloud IoT paradigm and a Li-Fi-based hybrid cloud framework study are described in [15]. In their study, a layered architecture comprising of infrastructure layer or the physical layer, local or private cloud layer and global or public cloud layer is proposed. The communication between the local cloud and the IoT devices in the physical is established using Li-Fi enabled access point which acts as a forwarding device. The IoT devices will be embedded with LED and photodiode which will make the light communication possible [12].

### *5.3 Li-Fi and Real-Time Analytics*

Time critical applications such as tracking and navigation are mainly based on real-time analytics. The large amount of data gathered is analyzed by applying logic and mathematical concepts to make conclusions, decisions and future predictions. The current business world which concentrates on making all the services much easier for customers keeps a constant watch on the data which can be the customer behavior pattern and gives suggestion for the next activity based on the analysis. This can be applied in the area of online shopping, medical sector, traffic system and so on. The Li-Fi-based companies have already started working with the real-time data analytics using the data which gets stored on their cloud using indoor positioning system (IPS). IPS is a system similar to global positioning system (GPS) but it is within the indoor environments where the satellite technologies are less efficient.

There are different technologies for IPS like acoustic systems, proximity based, infrared systems, Bluetooth based and so on. GeoLi-Fi is a Li-Fi-based geolocation platform by Li-Fi supporting companies like Basic6 and Oledcomm for indoor environments. Light fidelity access point installed in a dense network will gather accurate location information and create a map. The system comprises of Li-Fi LEDs that stream information as unique identifiers, LED Li-Fi modulator which transforms LED lights into a very simple Li-Fi antenna, Li-Fi-enabled tablets utilizing software communicate to consumers, centralized managerial interface and tools to perform the analytics, planning, forecasting and management. Each of the light source has a unique identifier and a mapping is defined of where all these lights are located and a unique identifier for those positions [16]. The tablets which are customized with the photodiode, and therefore whenever the tablet is under the Li-Fi enabled light its location is accurately identified. Instead of customized tablets, it is also possible to make use of Li-Fi-enabled dongle which can be easily plugged into our smartphones.

This system can be used in big supermarkets and the customer shopping pattern which is tracked using IPS is sent to cloud, real-time analytics can be applied, and thus the future shopping pattern can be predicted accurately. It can also make the shopping process much easier by giving information regarding the necessary item's location in the store and any additional information regarding it. This Li-Fi-based IPS can also be installed in emergency places like hospitals so that optimized navigation can be made without any delay [6].

## 6 Prototype Description

In this prototype, light is used as a medium of transmission. With the advent of Internet of vehicles, the entire traffic is going to be automated soon. Hence, using the headlights of the vehicles, an automated brake system can be implemented with Li-Fi support [2]. Earlier spread spectrum mechanisms were used in order to meet the needs of communication between vehicles. The military vehicle services are maintained and scrutinized based on the spread spectrum methodology.

A major drawback of the traditional system is that it requires the complete attention of the driver to control the speed and occurrence of a collision. But according to the proposed system in [2], the alert signal send to the vehicle as demonstrated with the help of light signals. Here, automation can be achieved as the distance between the vehicles if reduces below a specific limit and the controller reacts and hence avoid the accident. Basically, the light rays emitted by the car at the front through the brake light/stop light will be received by the car behind through a receiver at the front bumper which triggers the car according to the situation. As the emission is purely based on light rays, there is no chance of environmental issues or scattering of important signals.

## 6.1 Hardware Requirements

(a) Arduino UNO: The Arduino Uno is a microcontroller board based on the Microchip ATmega328P microcontroller and it is programmed with the help of an open-source Arduino software. This software acts as an integrated development environment (IDE) which will help to perform coding on the computer side and upload it to the physical board. Thus, we can say that Arduino UNO has hardware and software parts. The hardware section of the board comprises of digital and analog input/output (I/O) pins that may be interfaced to other circuits. This Arduino UNO is enough for the data transfer from one device to another.
(b) LED: Li-Fi bulbs are outfitted with a chip that modulates the light imperceptibly for optical data transmission. Light emitting diodes (LEDs) are used in Li-Fi as visible light transmitters. Li-Fi data is transmitted by the LED bulbs and received by photodiodes. The LED is connected to one of the digital pins. The LED blinks according to the binary logic sent from the processing software.
(c) LDR: An LDR is a component that has a (variable) resistance that changes with the light intensity that falls upon it. It is a passive component, which uses the concept of photoconductivity and has a resistor whose resistance value decreases when the intensity of light decreases. This optoelectronic device is mostly used in light varying sensor circuits. The light from the LED is identified by using LDR and the data is transmitted to the Arduino.
(d) Solar Panel 5v: In our system, it acts as a data receiver of analog digitals. Here, we have used 5v solar panel. In our system, solar panel acts as a data receiver of analog signals and not a power source. The LED light source combined with a solar panel can form a transmitter-receiver system. LED transmits encoded information which is received by the solar cell and made available where the information is required.
(e) Ultrasonic Sensor: An ultrasonic sensor is an electronic device that measures the distance of a target object by emitting ultrasonic sound waves and converts the reflected sound into a signal. In our system, the ultrasonic sensor is fixed in front of our vehicle and it measures the distance between the car and obstacle in front and gives an alerting signal if the distance is beyond a certain limit.
(f) Speaker and audio amplifier
(g) LCD display and batteries.

## 6.2 Software Requirements

(a) Arduino programming language: Arduino is an open-source platform used for building electronics projects. Arduino consists of both a physical programmable circuit board (often referred to as a microcontroller) and a piece of software, or integrated development environment (IDE) that runs on your computer, used to write and upload computer code to the physical board. The

IDE supports C and C++ and includes libraries for various hardware components, such as LEDs and switches. The program which is also known as sketch is uploaded to the Arduino board via a USB cable. Here, it can be run and will remain in memory until it is replaced.

(b) Python IDLE: Python IDLE is an integrated development environment (IDE) for Python. The Python installer for Windows contains the IDLE module by default. IDLE can be used to execute a single statement just like Python shell and also to create, modify, and execute Python scripts.

(c) Tinkercad: Tinkercad is a free, online 3D modeling program that runs in a Web browser, known for its simplicity and ease of use. We have used it to stimulate circuits and to generate Arduino codes.

## 7 System Design

The system uses the ultrasonic sensor which reads the distance to the obstacle in front of the vehicle where the sensor is fixed. The distance is then sent to the Arduino UNO. Basically, the light rays emitted by the car at the front through the brake light/stop light will be received by the car behind through a receiver at the front bumper which triggers the brake according to the situation. As the emission is purely based on light rays, there is no chance of environmental issues or scattering of important signals.

The prototype makes a study on digital and analog data transmission. For analog data transmission, we made use of audio source, LED, solar panel and audio amplifier.

### *7.1 Analog Data Transmission*

See Fig. 3.

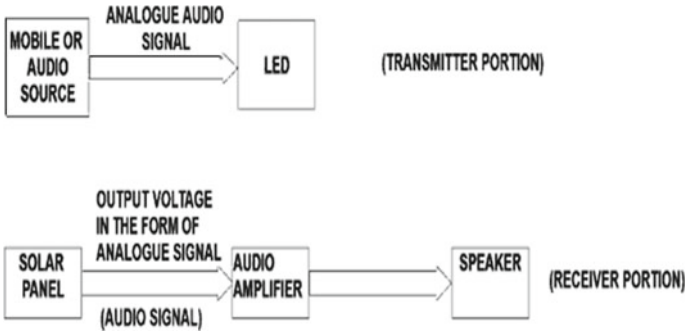



**Fig. 3** Block diagrams of analog data transmission

### *7.2 Schematic Diagram for Analog Data*

(a) Transmitter Section:

Depending upon the change in the music which is the analog signal the LED glows and its light intensity varies according to the fluctuations in the analog signal (Fig. 4).

(b) Receiver Section

At the receiver side, solar panel is used for light detection from the transmitter side and it then given to an audio amplifier followed by the speaker (Fig. 5).

## 8 Digital Data Transmission

The light signals emitted from the vehicle at the front will be received by the car behind via an LDR. Light-dependent resistors are very sensitive to light. When the required signal reaches the LDR, the input is also obtained. Ultimately when the distance between the vehicle or vehicle and the object is less than or equal to predefined limit known as safe distance, and the vehicle itself will apply brakes to

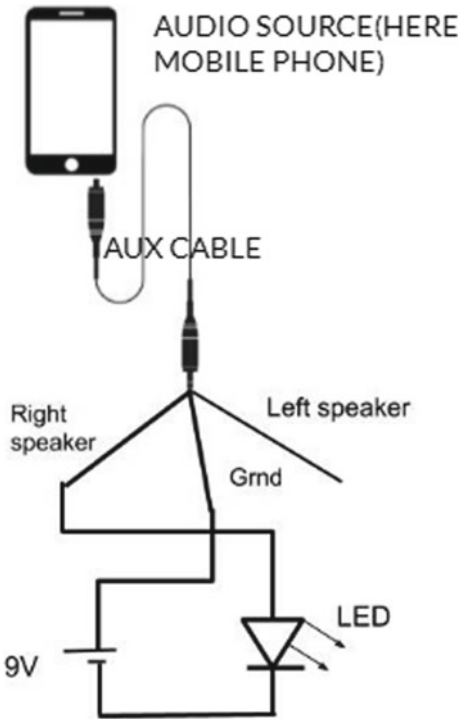



**Fig. 4** Block diagram of transmitter section

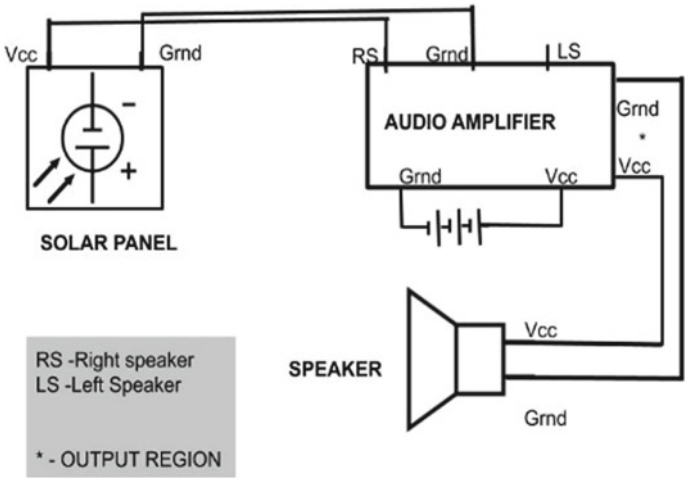



**Fig. 5** Block diagram of receiver section in analog data transmission

avoid the accident. If the distance between the vehicle are more than the defined safe distance, then the system will give a warning and emergency notification to the driver as a safety measure. The distance between the car and obstacle is measured by the ultrasonic sensor and then passed on to the Arduino UNO. Then, if it is less than the limit, then it is passed on to the LED and then the LED blinks accordingly.

## *8.1 Transmitter Section*

The value measured using ultrasonic sensor is stored into a variable, say 'time'. We then take the half of it and store it in a variable 't'. This value is used to compute the distance which is calculated as 340/t, as 340 m/s is the speed of the sound in air. The safe distance between car and the object is taken as 15 or above for our study. Here, a checking is done again to find if it is less than or equal to 5 and if so the LED glows for a time duration of 25 ms indicating brake should be applied otherwise the LED will glow for a time duration of 20 ms which is a warning sign. This is how the transmitter gets activated and starts sending data as Li-Fi signals and is depicted in the flowchart of the transmitter section (Fig. 6).


**Fig. 6** Block diagram of transmitter section in digital data transmission

## 8.2 Receiver Section

The LDR records data from the LED blink in 0 s and 1 s. It is then passed on to Arduino UNO and the alert or the warning message is sent. Here, instead of buzzer system, we used an LCD display. The value read from LDR is stored in a variable 'd'. In the next 20 ms, LDR value taken is stored to the variable val2. A string variable duration is concatenated with this val2. If val2 value is 1, then we perform a checking of whether duration is '0001' and if so it displays a warning in the LCD Display else if the duration is '00001' it causes the braking system to be enabled. The flowchart for the receiver section depicts this process (Figs. 7, 8 and 9).

The digital data transmission is represented as depicted in the flowchart (Figs. 10 and 11).

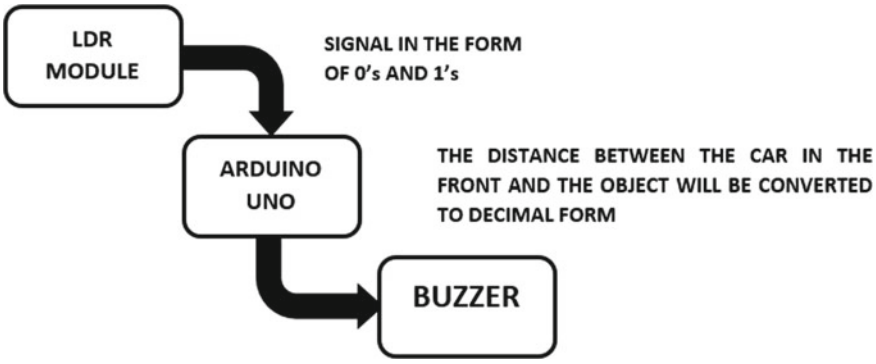



**Fig. 7** Block diagram of receiver section in digital data transmission

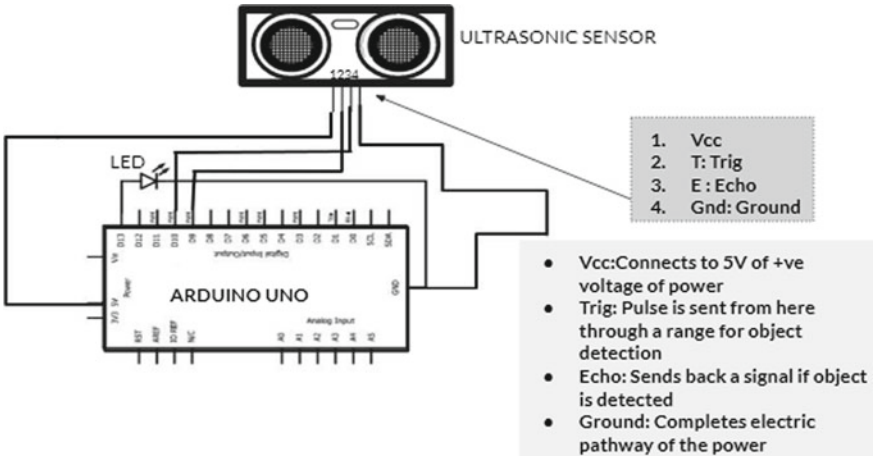



**Fig. 8** Schematic diagram of digital data transmission section

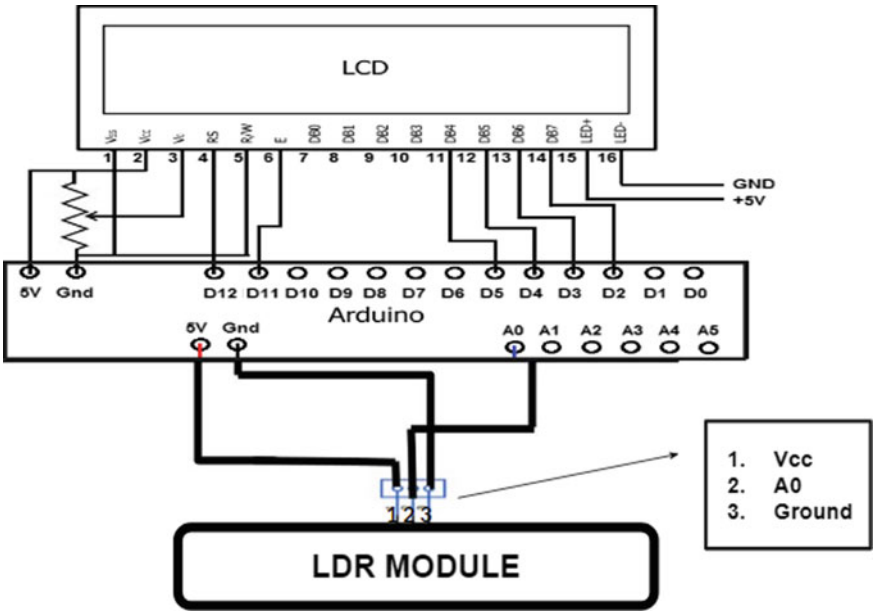



**Fig. 9** Schematic diagram of digital data receiver section

## 9 Results

In the system showing analog data transmission, we inferred that the speaker produces sound only when it is exposed to a light which is having Li-Fi signals. In the digital data transmission system, the audio signal from the mobile phone using AUX cable and was given to the Arduino. And this is the output waveform which we got from the Arduino. Now, it is clear that Li-Fi is an ideal point to point communication medium among two vehicles.

The plot that the signals have a voltage level variation depending upon the light signals. Similarly, the pattern was observed by varying the sensitivity of the LDR module to ensure the promptness of Li-Fi. We have voltage measured in the *y*-axis and the time taken in the *x*-axis.

The first plot shows output from solar panel when there are no variations (Figs. 12, 13 and 14).

## 10 Limitations and Challenges

1. Difficulty in calibrating the Li-Fi receiver in presence of ambient light.
2. The distance between the Li-Fi transmitter and receiver affects the visible light communication (VLC) technology.

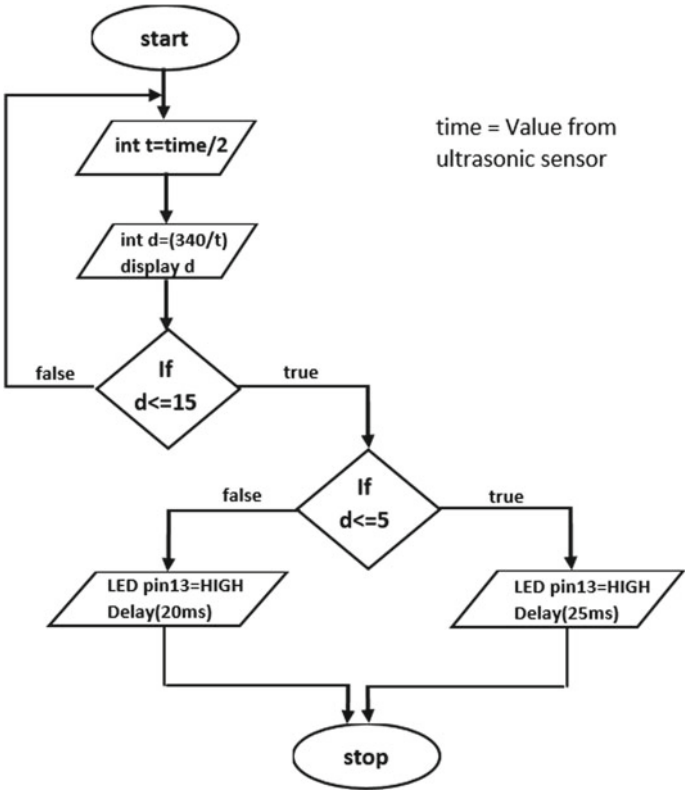



**Fig. 10** Flowchart of transmission section

3. In some cases, the shorter range of Li-Fi communication, Li-Fi put forwards security features.
4. We can use laser for long-range Li-Fi communication/data transfer, but in this case, the components of the setup should be in a fixed position.
5. Refraction cases Li-Fi signal loss
6. The modulation in Li-Fi is a challenge when the light illumination is low.
7. Li-Fi should need line of sight for effective data transmission. Small difference leads to interruption in the transmission.

## 11 Conclusion and Future Works

Li-Fi era is not too far and there are many companies who have started marketing products based on Li-Fi technologies. There are many countries where Wi-Fi usage is narrowed due to its harmful effects especially for children and health sectors. Li-Fi will predominantly gain wide spread popularity in those places because of its secure

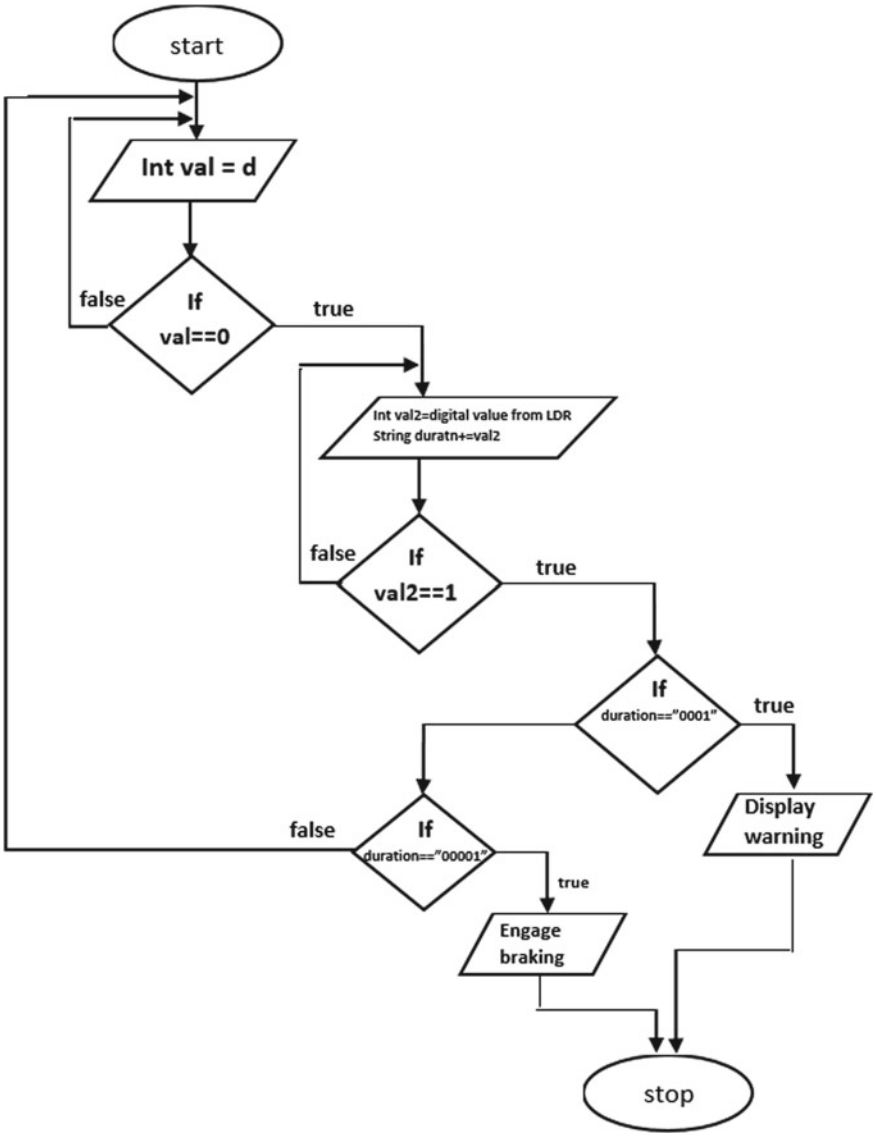



**Fig. 11** Flowchart of receiver section

and environment-friendly nature. It will gradually conquer the whole world and as predicted by the researchers working on it. As the wave made by photon produced by a light emitting diode has no power to penetrate through the skin it will soon be a safer substitute for the electromagnetic waves in the field of communication which can penetrate through even walls. In hospitals, where real-time health monitoring is

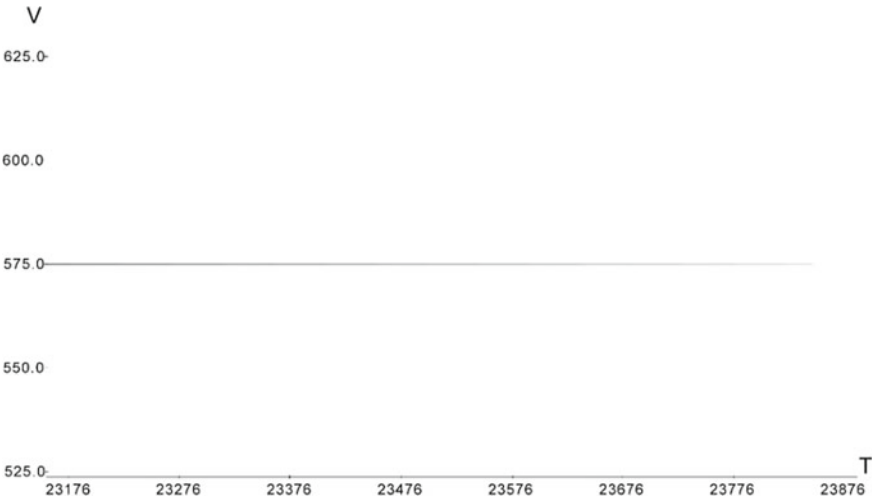



**Fig. 12** Graph showing the output from solar panel under a tube light (i.e., when there is no variation)

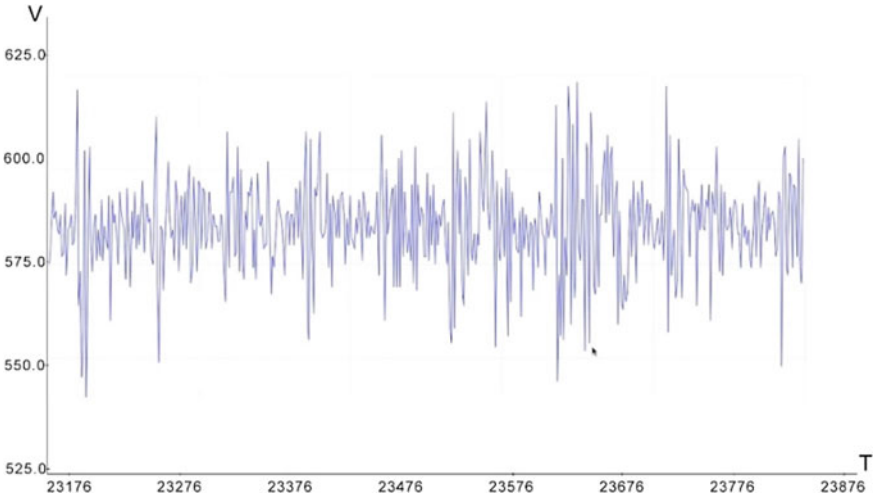



**Fig. 13** Graph showing the output from solar panel under a light integrated with audio signal

extremely time critical, Li-Fi can be used as it provides high speed. Li-Fi as elaborated in the applications can play predominant role in aircraft, underwater systems, vehicle to vehicle communications and other infrastructures. Even our street lamps can provide us with high speed Internet with the Li-Fi technology.

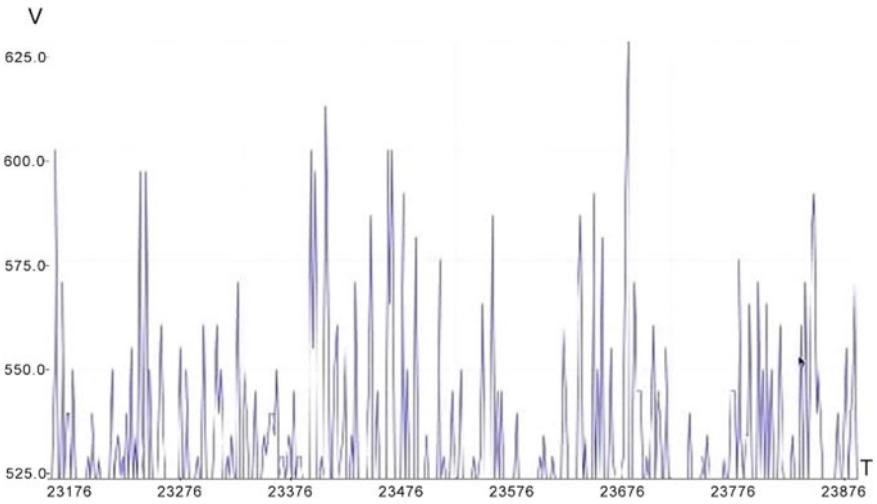



**Fig. 14** Graph showing the output from Arduino when audio signal from a mobile phone using AUX cable

## References


1. Ayyash M, Elgala H, Khreishah A, Jungnickel V, Little T, Shao S, Rahaim M, Schulz D, Hilt J, Freund R (2016) Coexistence of Wi-Fi and Li-Fi towards 5G: concepts, opportunities, and challenges. IEEE Commun Mag 54. https://doi.org/10.1109/mcom.2016.7402263
2. George R, Vaidyanathan S, Rajput A, Kaliyaperumal Da (2019) Li-Fi for vehicle to vehicle communication—a review. Procedia Comput Sci 165:25–31. https://doi.org/10.1016/j.procs.2020.01.066
3. Matheus LEM, Vieira AB, Vieira LFM, Vieira MAM, Gnawali O (2019) Visible light communication: concepts, applications and challenges. IEEE Commun Surv Tutorials 21(4):3204–3237, Fourthquarter. https://doi.org/10.1109/comst.2019.2913348
4. Ramadhani E, Mahardika GP (2018) The technology of Li-Fi: a brief introduction. In: IOP conference series: materials science and engineering 325(2018):012013. https://doi.org/10.1088/1757-899x/325/1/012013
5. Lokesh S, Priya N, Divyakanni K, Karthika S (2017) A Li-Fi based data transmission for anti-collision system. Int J Smart Sens Intell Syst 212–224. https://doi.org/10.21307/ijssis-2017-247
6. Yu HK, Kim JG (2019) Smart navigation with AI engine for Li-Fi based medical indoor environment. In: 2019 International conference on artificial intelligence in information and communication (ICAIIC), Okinawa, Japan, 2019, pp 195–199. https://doi.org/10.1109/icaiic.2019.8669041
7. Balaji K, Murugan SS (2019) Implementing IoT in underwater communication using Li-Fi. Int J Recent Technol Eng (IJRTE) 8(2S4). ISSN: 2277-3878
8. Hernandez-Oregon G, Rivero-Angeles ME, Chimal-Eguía JC, Campos-Fentanes A, Jimenez-Gallardo JG, Estevez-Alva UO, Juarez-Gonzalez O, Rosas-Calderon PO, Sandoval-Reyes S, Menchaca-Mendez R (2019) Performance analysis of V2V and V2I Li-Fi communication systems in traffic lights. Wirel Commun Mobile Comput 2019:12, Article ID 4279683. https://doi.org/10.1155/2019/4279683
9. Panhwar M, Khuhro SA, Mazhar T, Zhongliang Deng (2020) Li-Net: towards a smart Li-Fi vehicle network. Indian J Sci Technol 13:1–9. https://doi.org/10.17485/ijst/v13i18.210

10. Leba M, Riurean S, Lonica A (2017) Li-Fi—the path to a new way of communication. In: 2017 12th Iberian conference on information systems and technologies (CISTI), Lisbon, 2017, pp 1–6. https://doi.org/10.23919/cisti.2017.7975997
11. Kuppusamy P, Muthuraj S, Gopinath S (2016) Survey and challenges of Li-Fi with comparison of Wi-Fi. In: 2016 International conference on wireless communications, signal processing and networking (WiSPNET), Chennai, 2016, pp 896–899. https://doi.org/10.1109/wispnet.2016.7566262
12. Smys S, Basar A, Wang H (2020) Artificial neural network based power management for smart street lighting systems. J Artif Intell 2(01):42–52
13. Pottoo SN, Wani T, Dar A, Mir A (2018) IoT enabled by Li-Fi technology
14. Al-Ahmadi S, Maraqa O, Uysal M, Sait SM (2018) Multi-user visible light communications: state-of-the-art and future directions. IEEE Access 6:70555–70571. https://doi.org/10.1109/ACCESS.2018.2879885
15. Sharma P, Ryu J, Park K, Park J, Park J (2018) Li-Fi based on security cloud framework for future IT environment. Human-centric Comput Inf Sci 8. https://doi.org/10.1186/s13673-018-0146-5
16. Sungheetha A, Sharma R (2020) Cost effective energy-saving system in parking spots. J Electron 2(01):18–29

# Approaches in Assistive Technology: A Survey on Existing Assistive Wearable Technology for the Visually Impaired


**Lavanya Gupta, Neha Varma, Srishti Agrawal, Vipasha Verma, Nidhi Kalra, and Seemu Sharma**

**Abstract** People with visual impairment face a lot of challenges in their daily lives, be it small or big. This is mainly due to the lack of assistance provided by the modern assistive devices in terms of providing self-independence and the cost matching it. The main aim of this article is to research and explore the existing assistive technologies in the domain of visual impairment aid. The main objective of the assistive technology is to provide assistance in the day-to-day tasks, with a simple and wearable design to deliver a better user experience. This paper focuses on different approaches that will help the visually impaired through technology and learn those technologies that leverage a comfortable experience to the user. The primary objective of this survey is to navigate through the different approaches to find out the best suite for the authors in developing their technology.

**Keywords** Visually impaired · Assistive technology · Wearable device

## 1 Introduction

Human communication is at its root, where it is completely based on the text and speech [1]. While this communication has been accepted as a daily part of our life that requires little effort, it is not as easy for some. According to World Health Organization [WHO] report on vision [2–4], the number of visually impaired people present across the globe is approximately 2.2 billion, of whom at least 1 billion have a visual impairment not yet been addressed or prevented. According to a study published as of 2010 in the Global Estimates of Visual Impairment [5], almost 20.5% of the world's unsighted [6] are Indian citizens, along with 22.2% of the world's low vision population, and 21.9% of those are with vision impairment [7]. Disability takes a toll on

L. Gupta · N. Varma · S. Agrawal · V. Verma · N. Kalra (✉) · S. Sharma
Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India
e-mail: nidhi.kalra@thapar.edu

S. Sharma
e-mail: seemu.sharma@thapar.edu



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_42

a person's day-to-day life. With the ever-expanding and unparalleled technological growth, this part of the population will actually require a modern device that helps them to be more self-reliant and efficient without increasing the cost/expense [8]. Such modern devices are termed as 'assistive devices', and as the name suggests, they assist the user in partially or wholly for overcoming a problem/disability. Among all the other available assistive devices, wearable devices have been proven as the most efficient and helpful, out of which the head-mounted display [9] devices and clip-on devices remain as the most popular consideration. Recent studies and technologies in clip-on devices [10] have also shown promising results. A large fraction of the devices are available for assistance and are either focused on a singular aspect of the problem, or it is not efficient as expected [11].

This paper focuses on different approaches that are considered to aid the visually impaired through technology and compare the solutions with a new concept (focusing specifically on the user) to see how it is different, or in some cases, better than what is already available in the market.

## *1.1 Types of Computer Glasses*

Different types of computer glasses exist, and this adds information alongside what the user can see, or in this case, it will not be able to view. Like other computers, smart glasses may collect information from external or internal sensors. Using the sensors, smart glasses can guide the wearer, inform the user about his/her surroundings, warn them of threats and generally assist them in carrying out simple tasks that would otherwise be difficult for them [12, 13]. Various types of computer glasses are mentioned below.

### 1.1.1 Monocular Smart Glasses [14]

Monocular smart glasses as the name suggests are glasses in the form of HMDs where an optical engine is placed on one of the lenses [15, 16]. Vuzix M300, Google Glass [17], Optivent's Ora-2 and the Lumus Sleek are some examples of such devices.

### 1.1.2 Binocular Smart Glasses [18]

Binocular smart glasses are another form of a head-mounted display system which consists of two transparent displays, which deliver a stereoscopic view to the users. The advantage of such HMDs over monocular glasses is that a more diverse enhancement of the overall view perspective of the wearer is conceivable because of the optical engines mounted before each eye. Much like monocular glasses, these also display the information just beyond the user's line of sight, but to both eyes. Verge

IT [19], Epson Moverio [20], the ODG R-9 [21], the Sony SED-E1 [22, 23] and SED-100A [24] are some of the current binocular smart glasses in the market.

### 1.1.3 Audio Augmented Reality Smart Glasses [25]

ARSG are basically head-up transparent displays that add virtual information to what the user sees with an integrated wearable miniature computer [26]. As the name suggests, the glasses provide the users with augmented reality by overlaying virtual data on the real worldview of the wearer, and doing so makes it conceivable to enhance a human's perception of reality [27]. ARSG are an ideal user interface for industry usage by workers since they are hands-free gadgets that show information at eye level, right where it is required.

### 1.1.4 Immersive or Mixed Reality Smart Glasses [28]

Mixed reality is the combination or merger of the real-world view with the virtual world to create new environment. A real-world environment is superimposed with virtual objects with the help of computer graphics. The objects visually act like a physical object in the real-world environment, change shape, and appearance according to the angle and position of where they are interacted from and also, based on the external environment or actions/movements of the user. Hence, a user can view the virtual object from different bearings and directions by moving around spatially in a room, as if it were a real object [29].

Some mixed reality glasses are Microsoft HoloLens, ODG-9 AR and VR glasses and the Meta Space glasses. These gadgets are fully immersive and inclusive standalone systems with outstanding displays, allowing you to render 3D objects on board [30].

A field that has not been utilised to the best of its capacity is that of **audio augmented reality** [31]. Audio augmented reality is when the real or physical world can be augmented with a virtual layer of auditory signals, which allows the user to acquiescently interact with it. In certain glasses, the sensors on the glasses gather appropriate information and the focused speaker directly informs the user. The technology in this field can be used to provide maximum utility to the user.

## *1.2 Motivation for Conducting the Survey*

The motivation for this survey was to clearly analyse the existing technologies/devices that exist, which aid the visually impaired and the features incorporated in them. This analysis helps to identify the problems that were solved by these technologies and the ones that could not be. Along with that, it provides a

detailed overview of how these technologies can be better utilised to serve the visually impaired and make it easier for them to participate in their day-to-day activities. An exhaustive survey was conducted, particularly motivated from these facts:

- The need and demand for understanding the benefits and drawbacks of the impact of existing assistive devices for the visually impaired.
- The need and demand for understanding the various technologies these devices use.
- Can these technologies be utilised in a better way to suit the needs of the customer base?
- Can newer technologies be incorporated in such devices to provide an enhanced product?

Thus, there is an immediate need to understand and explore technologies that can be used to assist the visually impaired and the status of their applications in the physical world. Not only does the technology assist the visually impaired, but its application also includes safety; thus, it is our duty to work towards the enhancement of this area, so as to make it a better place to live in for those with disabilities. However, to gain an understanding of technologies that can be used, it is highly required to analyse the use of existing technologies and their applications, and only then it can contribute towards making a change in this field.

### *1.3 Outline*

The remaining article is designed as: Sect. 2 presents the background. Section 3 goes into a detailed analysis of the existing devices and technologies, highlighting their use in the device and their impact. Section 4 highlights the inferences made by conducting researches on the existing devices and what better technologies can be used in those devices. Section 5 presents the concept of a prototype device that could be used to effectively solve various problems faced by the visually impaired population, in a single device. The proposed research work is concluded in Sect. 6.

## 2 Background

Until recently, solely medical solutions and practices used for treating disabilities have become more popular. Any assistive technology will only be considered as a branch of that medical practice. It is only in the past few years that assistive devices have really taken the centre stage in not just aiding people with visual disabilities but also making them largely autonomous and independent [32].

Particularly, the problems like navigation, reading and providing visibility to people with partial visibility problems are considered as the different challenges associated with the technologies.

Society today has changed considerably to eradicate the discrimination against impaired population. The combination of new laws and awareness has asserted the idea of acceptance and independence of the impaired. But with that, there is a need to devise a highly effective way to help the visually impaired in order to move a step closer towards normalcy by asserting self-dependence. This can only be done by implementing various advanced technologies. Various products have been integrated in the past few years to help the visually impaired to deal with these problems. This article will extensively explore these products along with the technologies employed and how they evolved over the years.

## *2.1 Review Process*

Figure 1 shows the growth of research in the field of assistive technologies for the visually impaired since 1990 until date. The graph shows a tremendous increase in research contributions in the last decade (Table 1).

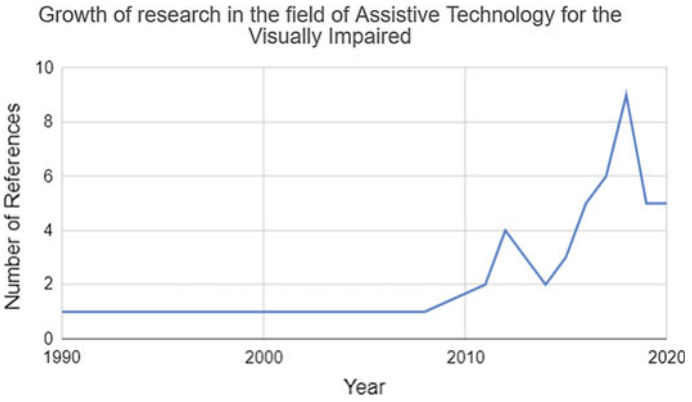



**Fig. 1** Graphical representation of the growth of research in the field of assistive technology for the visually impaired

**Table 1** Classification of papers

| Property | Categories |
|---|---|
| Year | 1990–2020 |
| Methods | Ultrasound-assisted obstacle detection [33], human assistance, sonar sensors, visual sensors |
| Targeted problems | Travel/navigation, reading [34], enhancing vision [35] |
| Publication type | Research paper, journal article, conference proceedings, Web article, magazine article |

Figure 2 shows a graphical representation of the distribution of research articles in various sources such as journals, conferences and proceedings. The graph depicts the maximum contribution of research articles in journals followed by Web articles.

The proposed research work has framed 6 research questions as shown in Table 2. These research questions helped the authors to explore the literature and collect all the information related to assistive technologies for the visually impaired.

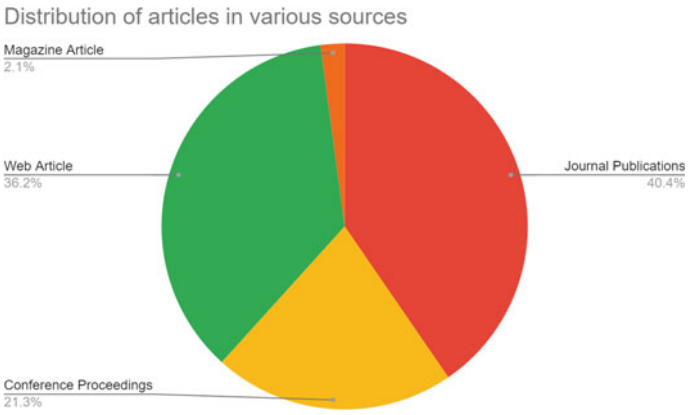



**Fig. 2** Graphical representation of the distribution of articles in various sources, used for the review process

**Table 2** Research questions

| Research questions | Main motivation | Section referred |
|---|---|---|
| *RQ1.* What is assistive technology? | Identify the meaning of assistive technologies and to understand the current and existing technologies | Sect. 1 |
| *RQ2.* What is the need for such technologies? | Identify the needs of the people using such technologies. Assessing existing technologies features with respect to the needs of the people identified | Sect. 1.2 |
| *RQ3.* What are the different types of assistive technology being used? | Identify the different types of assistive technology | Sect. 1.1 |
| *RQ4.* How has assistive technology evolved? | Identify the evolution of assistive technology and finding the latest trends | Sect. 3 |
| *RQ5.* How helpful is the current technology in aiding the user? | Identify the pros and imperfections of the current assistive technology | Sect. 4 |
| *RQ6.* Can the loopholes in current technology be overcome by a new device? | Identify what better technology can be used to reduce the cost and increase the effectiveness of devices | Sect. 5 |

## 3 Related Works

### 3.1 BuzzClip [36]

BuzzClip caters to that part of the society that lives with blindness or partial sight. It aims to reduce the dread and anxiety related to navigation by providing a dependable way to make the user aware of their immediate surroundings and obstacles.

Construction: BuzzClip is a small and discreet wearable device. It aids in navigation by detecting obstructions in the wearer's path using ultrasound. It is an additional supportive device to the cane and can also be used on its own for those with partial sight looking for a subtle solution.

Methodology: It informs the user of the hindrances through instinctive vibrations and helps them safely navigate their way around these objects. It additionally offers fundamental obstruction detection at head level and can be attached to clothing, proving it exceptionally adaptable and helpful.

### 3.2 Aira AR Smart Glasses [37]

The Aira smart glasses, developed by La Jolla, are augmented reality smart glasses that are intended to aid users who are completely or partially visibly impaired, to read significant text (expiration date on grocery items, information on medicines, etc). It is the blend of cutting-edge technology, hardware and live assistance support provided by highly attentive aides.

Construction: The glasses are equipped with a camera and inbuilt earpiece.

Methodology: With these smart glasses, Aira has built a service that fundamentally places a human assistant into a visually impaired user's ear and sends live video of the surroundings of the user through the camera on the glasses to the assistants, who can then give instructions to the end-users and help them with their basic tasks, be it directions or describing scenes.

### 3.3 Sunu Band [38]

Sunu manufactures the Boston-based wearable, one of which is Sunu Band: a sonar sensor device that helps the visually impaired not collide with people and objects by guiding them.

Construction: Uses sonar sensors for obstacle detection portable system [39].

Methodology: When the sensor detects an object or person 15 feet away from the user, the device vibrates to let the user know of a potential collision, and with increasing proximity, the device vibrates with increased frequency and intensity.

### *3.4 Maptic [40]*

Maptic devices are wearable navigation systems.

Construction: A necklace that acts as a visual sensor and a device that clips around the wrist of the user acts as a series of feedback units. A voice-controlled iPhone app uses GPS to direct the user.

Methodology: The device vibrates to guide the user to desired destinations. To steer the user towards either direction, the device sets off a series of vibrations to either side of the body.

### *3.5 MyEye2 [41]*

The aim of these low-vision electronic glasses is to make day-to-day activities like reading, writing, recognising faces and items easier for the visually impaired.

Construction: Using a magnetic mount, the device is attached to any pair of glasses. A light camera is attached to the sides of the glasses' frame as well.

Methodology: The power button of the OrCam device is near the speaker, which is located on the flat underside. A touch bar (raised for easy location) is placed on the top side of the device which is rounded. It allows the user to read multiple languages, identify currency and also face recognition. OrCam also learns to identify objects, as it does with faces.

### *3.6 NuEyes Pro [42]*

NuEyes Pro is a pair of head-mounted smart glasses that are lightweight and wireless. It is specially designed to aid the visually challenged to see better. NuEyes Pro can help with conditions like glaucoma, diabetic retinopathy and macular degeneration. It converts print to speech and allows users to carry out daily activities like watching TV, writing and reading.

Construction: An image is captured by the camera placed at the front of the glasses and then shown in a magnified form inside of the lenses. Magnification of up to 12

times can be obtained on images, which helps the user see the finer details more clearly. NuEyes also has a 30° field of view (FOV) and a 1080i stereo display.

Methodology: A remote-handheld controller and vocal orders are used to control the device.

### *3.7 OXSIGHT [43]*

In 2016, OXSIGHT was established based on the effort of a team of Oxford University, who started researching and working closely with the low-vision community a decade ago. Their main aim was to bring out an innovative device to help this community. Its main features include expanding field of vision, enhancing light and shape detection. At Oxford University, Dr Stephen Hicks and his colleagues researched and came up with the idea of these smart assistive glasses that would be able to present essential information about the proximity of obstacles from the wearers. This was achieved with the help of a camera and display option. Along with that, the glasses are also capable of increasing the brightness and clarity of obstacles.

### *3.8 Intelligent Reader for the Visually Handicapped [1]*

This Raspberry Pi-based reader captures an image of some textual material and enables visually challenged users to passively interpret the text present in that image.

Technologies used: Raspberry Pi Camera [44], to capture the image; ImageMagick software is used to get a scanned image from the picture obtained; Tesseract Optical Character Recognition software obtains text converted from the image; Text to Speech (TTS) engine is used for the transfiguration of text into speech.

Table 3 effectively summarises these latest technologies and works in the field of visual aid devices and provides an efficient comparison between the same.

## 4 Discussion

The existing technology has greatly helped aid the visually impaired in providing them independence in day-to-day activities. They have immensely impacted their social status and mental well-being. While the contribution of these devices in bringing technological advancement (as in present day) to help the visually impaired is significant, there is always scope for improvement. Most of the existing wearable devices are not effectively solving the root of the problem. They focus only on subsets rather than the main obstacle. The most popular devices for smart glasses are Braille readers [45] which can read and write using the arrangement of dots.

**Table 3** Existing technologies to aid the visually impaired

| Device | Proposed technology | Year proposed |
|---|---|---|
| BuzzClip | Wearable device that aids the visually impaired by informing the user of hindrances in their path through instinctive vibrations and helps them safely navigate their way around these objects | 2018 |
| Aira AR smart glasses | Augmented reality smart glasses that are intended to aid users to read significant text. This is achieved by providing live assistance through human aid into a visually impaired user's ear and sending live video of the user's surroundings through the camera on the glasses to the assistants, who can then give instructions to the end-users and help them with their basic tasks | 2018 |
| Sunu Band | A sonar sensor device that helps the visually impaired not collide with people and objects by guiding them | 2018 |
| Maptic | A tactile assistive navigation device that works like a cane, without the stigma. It uses a necklace that acts like a visual sensor, a voice-controlled app on the phone and vibrations to help the user to navigate | 2018 |
| MyEye2 | These are low vision (partial visual impairment) electronic glasses that help make day-to-day activities like reading, writing, recognising faces and items easier for the visually impaired | 2017 |
| NuEyes Pro | Head-mounted smart glasses designed to aid the visually challenged to see better. NuEyes Pro can help with conditions like glaucoma, diabetic retinopathy and macular degeneration. It converts print to speech and allows users to carry out daily activities | 2018 |
| OXSIGHT | The aim of this device was to improve low vision. Features include expanding field of vision, enhancing light and shape detection. Along with that, the glasses are also capable of increasing the brightness and clarity of obstacles | 2016 |

(continued)

**Table 3** (continued)

| Device | Proposed technology | Year proposed |
|---|---|---|
| Intelligent reader for the visually handicapped | This Raspberry Pi-based reader captures an image of some textual material and uses text recognition algorithms to identify texts in those images and provide the user with an audio cue (using text to speech) | 2020 |

Another device commonly used is an audio reader which reads out books and papers saved in the audio format. This would essentially point to the fact that for each task, there exists an individual product solving only one part of the problem. Both these devices being costly, it leaves a huge dent on the pocket of those who really need them, especially those who will need to buy different assistive devices for different tasks. This ultimately leads to sacrificing basic comfort in their day-to-day lives.

Figure 3 gives a summary of the problems current assistive technology is solving. With the help of this figure and further research, one can conclude that there are only a few devices (35–44) that solve a multi-purpose function.

Moreover, most of the devices explored in this article focus on improving visibility for those who are only partially visible. The devices that claim to provide aid to all visually impaired, like Aira smart glasses, do not provide autonomy to the user and force them to lean onto a person that is aiding them through a live call feature. These techniques are not just outdated, but also unattractive to the visually impaired.

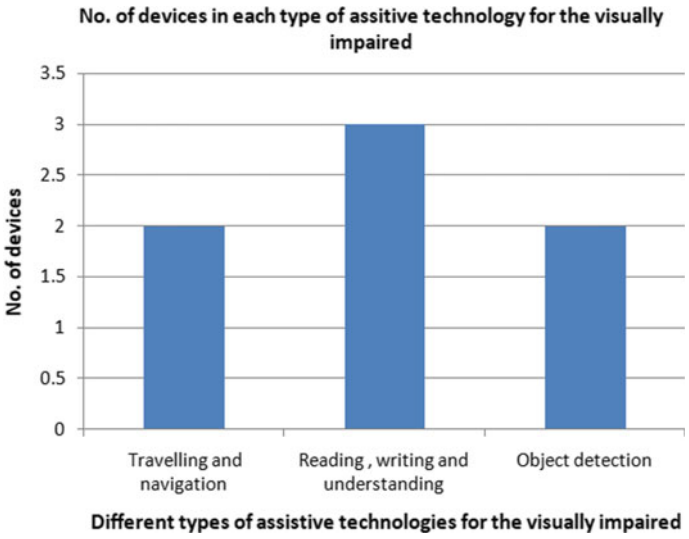



**Fig. 3** Graphical representation of the number of assistive technology devices for the visually impaired in each type of field

A few devices target text reading, others target object detection in the path of the user, and a minute number of devices help in recognising objects or faces using machine learning or artificial intelligence tools, but these are limited to partial visual impairment. And to date, no device exists that combines all the features into a single device. There has been no development of even a costly all-round device, let alone a cost-effective solution.

The technology is assistive devices started with object detection solutions using ultrasound and moved onto braille readers, and even after leaps of technological advancement, it seems that assistive devices have only progressed to improve or build upon the technology of existing object detection devices and text readers. Assistance in the form of recognition of faces or objects is largely absent. Artificial intelligence and machine learning have rarely been incorporated into these devices, and when they have been, the devices do not seem to include the most necessary features of navigation and obstruction detection.

Among existing features, there can be multitudes of advancement. Technologies like IoT can be used for wireless detection (similar to how it might be effective in vehicle collision detection [46]) instead of ultrasound. For live video assistance technologies, instead of limiting to verbal assistance by a human aid, motor control systems for wireless automation [47] can be explored to help navigate the user.

While object and face recognition is trivial in these devices, recognition models like R-CNN, fast R-CNN (and YOLO—you only look once) should be implemented for faster results, as real-time recognition requires speed.

Most importantly, there is still scope for technology that can combine various factors into a single device and provide immense autonomy to the visually impaired for most activities.

Research and studies have found out that as the number of assistive products and related technology increase, so does product suspension. When asked, visually impaired users of such products communicated disapproval of most assistive devices currently used because of the factors of tactile feeling and visual balance. Many complain of self-consciousness linked with the design of the device and its functionality. There has yet to be a breakthrough that can model a device according to the user's feedback and advice.

## 5 Device Concept

The main aim is to investigate the dynamics of converging mobile devices and their accessibility features with wearable devices coming into the market. This shall make the lives of visually impaired more comfortable and closer to that of the people around them.

In addition to that, the main focus is given to the possibility of developing a new technology, which would be able to accommodate the most crucial aspects of the problems faced by these people as well as manage to stay efficient by keeping its cost low.

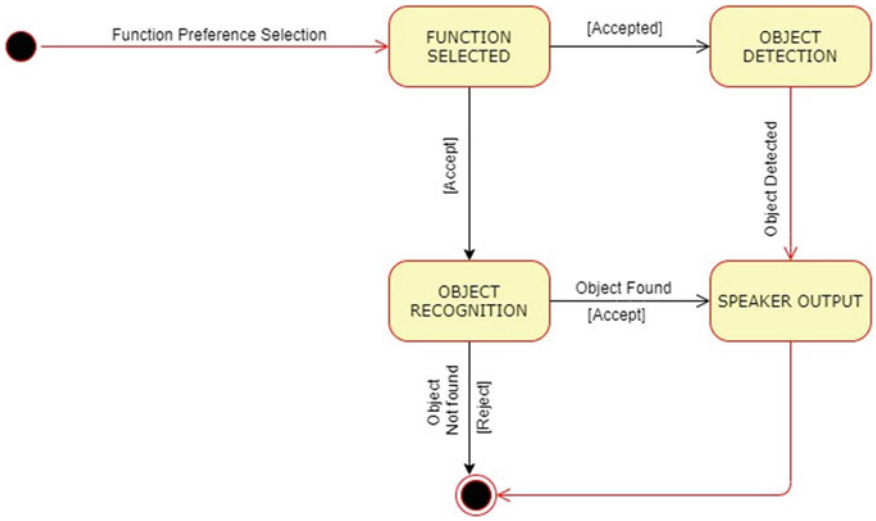



**Fig. 4** A general state chart representation of the proposed ideal assisting device

Figure 4 represents a state chart (flow chart) of the working of a close to ideal assistive device. Current devices only include the object detection flow path or a branch of the object recognition flow path (as text recognition).

## 6 Conclusion

The proposed research work leverages a new concept of smart device that aims to study the existing assistive technologies that were developed, to analyse their drawbacks, and furthermore it creates a technology that overcomes the challenges that were previously faced. While performing a detailed study, many research gaps are observed in the existing designs of this field. Most of these technologies focus only on one aspect of the problem; for example, the majority of the devices work on assisting the visually impaired with reading, or to an extent, identifying limited objects. With this device, it will always provide a comprehensive product that not only assists in their day-to-day lives but gives a wholesome experience to the user which they can enjoy without being constantly reminded of their disabilities. Most of the devices discussed above are focusing on just a fraction of the problem, making it difficult for users to have a better experience. This research work can be concluded with some findings that are listed below:

(1) Larger parts of the technologies already present in the market do not highlight all the problems and instead develop a product focusing on only one aspect of the issue.

(2) After finishing our research, it has been realised that out of all the technologies that has been analysed, where it cannot find any device, which addresses all the problems in one device.
(3) All the devices mentioned are very heavy on a middle-class person's pocket in comparison with the features it provides, making it very difficult to afford. This makes it really difficult for the disabled to even participate in day-to-day activities without proper assistance, and as a result of this situation it makes them feel dependent and liable.

The emerging innovative perspectives of different existing technologies can be incorporated into the making of new wearable technology. This work served the purpose of gaining insight into designing a new technology for smart devices which strives for maximum potential associated with minimum cost [48].

## References

1. Mainkar VV, Bagayatkar TU, Shetye SK, Tamhankar HR, Jadhav RG, Tendolkar RS (2020) Raspberry Pi based intelligent reader for visually impaired persons. In: 2020 2nd international conference on innovative mechanisms for industry applications (ICIMIA). IEEE, pp 323–326
2. Geneva: World Health Organization. World Report on Vision. https://www.who.int/publications/i/item/world-report-on-vision
3. Geneva: World Health Organization and the United Nations Children's Fund (UNICEF). A vision for primary health care in the 21st century: towards universal health coverage and the sustainable development goals. https://www.who.int/docs/default-source/primary-health/vision.pdf
4. Armstrong KL, Jovic M, Vo-Phuoc JL, Thorpe JG, Doolan BL (2012) The global cost of eliminating avoidable blindness. Indian J Ophthalmol 60(5):475
5. Pascolini D, Mariotti SP (2012) Global estimates of visual impairment: 2010. Br J Ophthalmol 96(5):614–618
6. Garewal NS India home to 20% of world's visually impaired. https://www.tribuneindia.com/news/archive/nation/india-home-to-20-of-world-s-visually-impaired-738167
7. Cimarolli VR, Boerner K, Brennan-Ing M, Reinhardt JP, Horowitz A (2012) Challenges faced by older adults with vision loss: a qualitative study with implications for rehabilitation. Clin Rehabil 26:748–757
8. Uslan MM (1992) Barriers to acquiring assistive technology: cost and lack of information. J Vis Impairment Blindness 86(9):402–407
9. Everingham MR, Thomas BT, Troscianko T (1998) Head-mounted mobility aid for low vision using scene classification techniques. Int J Virtual Reality 3(4):1–10
10. Device mounted on glasses for the blind provides a new way of independence. https://www.orcam.com/en/article/new-lease-sight-glasses-blind-see/
11. Avila M, Kubitza T (2015) Assistive wearable technology for visually impaired. In: Proceedings of the 17th international conference on human-computer interaction with mobile devices and services adjunct, pp 940–943
12. Hub SG What types of smart glasses are there? https://smartglasseshub.com/types-of-smart-glasses/
13. Wei J (2014) How wearables intersect with the cloud and the Internet of Things: considerations for the developers of wearables. IEEE Consum Electron Mag 3(3):53–56
14. Koneva TA, Romanova GE (2018) Designing of a monocular see-through smart glass imaging system. In: Digital optics for immersive displays, vol 10676. International Society for Optics and Photonics, p 106760v

15. Schlosser P, Grundgeiger T, Happel O (2018) Multiple patient monitoring in the operating room using a head-mounted display. In: Extended abstracts of the 2018 CHI conference on human factors in computing systems, pp 1–6
16. Russey C These new generation of wearables are empowering blind and the visually impaired. https://www.wearable-technologies.com/2018/12/these-new-generation-of-wearables-are-empowering-blind-and-the-visually-impaired/
17. Lv Z, Feng L, Li H, Feng S (2014) Hand-free motion interaction on Google glass. In: SIGGRAPH Asia 2014 mobile graphics and interactive applications, pp 1–1
18. Terhoeven J, Wischniewski S (2017) How to evaluate the usability of smart devices as conceivable work assistance: a systematic review. Advances In ergonomic design of systems. Products and processes. Springer, Berlin, Heidelberg, pp 261–274
19. Ahn S, Son J, Lee S, Lee G (2020) Verge-It: Gaze interaction for a binocular head-worn display using modulated disparity vergence eye movement. In: Extended abstracts of the 2020 CHI conference on human factors in computing systems, pp 1–7
20. Matei O, Vlad I, Heb R, Moga A, Szika O, Cosma O (2016) Comparison of various epson smart glasses in terms of real functionality and capabilities. Carpathian J Electr Eng 10(1)
21. Segan S These AR glasses are the first qualcomm snapdragon 835 products. https://in.pcmag.com/chipsets-processors/111195/these-ar-glasses-are-the-first-qualcomm-snapdragon-835-Products
22. Qiu X, Keerthi A, Kotake T, Gokhale A (2019) A monocular vision-based obstacle avoidance Android/Linux middleware for the visually impaired. In: Proceedings of the 20th international middleware conference demos and posters, pp 25–26
23. Develop Customized AR solutions. https://developer.sony.com/develop/smarteyeglass-sed-e1/
24. Add clarity to augmented reality solutions. https://developer.sony.com/develop/sed-100a-holographic-waveguide-display/specifications
25. Syberfeldt A, Danielsson O, Gustavsson P (2017) Augmented reality smart glasses in the smart factory: product evaluation guidelines and review of available products. IEEE Access 5:9118–9130
26. Barfield W (2015) Fundamentals of wearable computers and augmented reality. CRC Press
27. Kipper G, Rampolla J (2012) Augmented reality: an emerging technologies guide to AR. Elsevier
28. Sorko SR, Trattner C, Komar J (2020) Implementing AR/MR—learning factories as protected learning space to rise the acceptance for mixed and augmented reality devices in production. Procedia Manuf 45:367–372
29. Bottino AG, García AM, Occhipinti E (2017) Holomuseum: a prototype of interactive exhibition with mixed reality glasses Hololens. Doctoral dissertation, Master dissertation, Polytechnic University of Valencia
30. Yuan M, Khan IR, Farbiz F, Niswar A, Huang Z (2011) A mixed reality system for virtual glasses try-on. In: Proceedings of the 10th international conference on virtual reality continuum and its applications in industry, pp 363–366
31. Mynatt ED, Back M, Want R, Frederick R (1997) Audio aura: light-weight audio augmented reality. In: Proceedings of the 10th annual ACM symposium on user interface software and technology, pp 211–212
32. Hersh MA, Johnson MA, Keating D (2008) Assistive technology for visually impaired and blind people. Springer, London
33. Patel P Israeli startup's vision device can help the nearly-blind read and recognize faces. https://spectrum.ieee.org/the-human-os/biomedical/devices/israeli-startups-vision-device-can-help-nearlyblind-read-and-recognize-faces-study-shows
34. Merino-Gracia C, Lenc K, Mirmehdi M (2011) A head-mounted device for recognizing text in natural scenes. In: International workshop on camera-based document analysis and recognition. Springer, Berlin, Heidelberg, pp 29–41
35. Peli E, Lee E, Trempe CL, Buzney S (1994) Image enhancement for the visually impaired: the effects of enhancement on face recognition. JOSA A 11(7):1929–1939
36. The all new Buzzclip. https://www.imerciv.com/

37. Nguyen BJ, Kim Y, Park K, Chen AJ, Chen S, Van Fossan D, Chao DL (2018) Improvement in patient-reported quality of life outcomes in severely visually impaired individuals using the Aira assistive technology system. Transl Vis Sci Technol 7(5):30–30
38. Sunu, Inc. Sunu Band. https://www.sunu.com/en/index.html
39. Lesecq S (2018) Obstacle detection portable system: why do we integrate A UWB RF radar? Application to a smart white cane for VIB people. In: Workshop on emerging electromagnetic and RF systems for health monitoring and therapy, Stanford University
40. Farrington-Arnas E Maptic. https://emilios.co.uk/portfolio-maptic.html
41. Pauls J An evaluation of Orcam Myeye 2.0. https://www.afb.org/aw/19/8/15066
42. Hagen M Nueyes Pro-Smartglasses for low vision. https://www.closingthegap.com/nueyes-pro-smartglasses-low-vision/#:~:text=WhatisNueyes?
43. OXSIGHT products. https://www.oxsight.co.uk/
44. Camera module. https://www.raspberrypi.org/documentation/usage/camera/
45. Braillenote Touch 32 Plus—Braille Note Taker/Tablet. https://store.humanware.com/europe/blindness-braillenote-touch-plus-32.html
46. Dhaya R, Kanthavel R (2020) A wireless collision detection on transmission poles through IoT technology. J Trends Comput Sci Smart Technol (TCSST) 2(03):165–172
47. Vinothkanna R (2020) Design and analysis of motor control system for wireless automation. J Electron 2(03):162–167
48. Schreier EM (1990) The future of access technology for blind and visually impaired people. J Vis Impairment Blindness 84(10):520–523

# Stateless Key Management Scheme for Proxy-Based Encrypted Databases


**Kurra Mallaiah, Rishi Kumar Gandhi, and S. Ramachandram**

**Abstract** To preserve the confidentiality of important data stored in third-party managed platforms like public cloud databases is continuously raising security concerns. Such databases need to be shielded from malicious administrators or malicious software attacks, so as to increase the trust of the potential customers of such security. Therefore, it is very important that the data must be protected at rest, in transition and also while in operation to achieve full confidentiality for the data stored in databases. Shielding the confidentiality of data at rest and in transition is adequately addressed but shielding the confidentiality of data while in operation is still a big challenge. CryptDB provides confidentiality for relational databases by supporting the computations on encrypted data. The keys which are used in CryptDB are stored in the proxy. This paper proposes a stateless key management scheme, inversely the statefull key management scheme of CryptDB. The security proof is presented along with cryptographic security definitions and that the security of our key management under the random oracle model security assumption. The proposed key management scheme is the first of its kind that is a stateless key management. The scheme eliminates the storing of user cryptographic keys in the proxy and, hence, avoids possible attacks and analysis on keys in the proxy and also various other problems such as key backup, key loss, and key audit. Our proposed solution also satisfies regulatory compliance by not storing the cryptographic keys anywhere.

**Keywords** Trusted proxy · Security · Stateless key management · Database · Cloud computing

## 1 Introduction

In context of cryptographic solutions, key management refers to protection, storage, backup, and administration of encryption keys. The confidentiality of critical data stored in the cloud databases should be maintained against malicious administrator and malicious software attacks. While using these services managed by third parties,

K. Mallaiah (✉) · R. K. Gandhi · S. Ramachandram
University College of Engineering, OU, Hyderabad, Telangana, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_43

the physical control over data moves beyond the premises of user's organization into that of third-party administrator. The encryption techniques ensure the confidentiality of data, if and only if, encryption/decryption keys are accessible only to authorized entities. Therefore, the cryptographic keys used for encryption of data should be protected from malicious administrators or software hackers. The problem is that a single enterprise may end up using various different and maybe incompatible algorithms to protect confidential data. Subsequently, a vast number of cryptographic keys are generated which are in turn needed to be protected and used efficiently. Besides, separation of keys and data storage is challenging task. For example, the server compromised by an attacker where cryptographic keys along with encrypted data are stored, and then confidentiality of sensitive data becomes a bigger threat. Therefore, the keys along with sensitive data never be stored together. User or user's authorized entities should manage keys but not third-party administrator. So, the key management has become critical research in cloud computing. According to Verizon [15], about 42% companies have trouble in implementing a proper key management policy to protect data. Experts say, appropriate key management is becoming more important than encryption itself. So, keys speak for '*the keys to the kingdom,*' if anyone access to the cryptographic key, they may access to the most important data in the company. Appropriate key management is a requirement for compliance to PCI-DSS [16]. Even auditors scrutinize how companies govern cryptographic keys [5].

Nowadays, organizations are migrating company databases to cloud data centers rapidly. In cloud service, cloud provider controls IT infrastructure, and clients have to trust them [26]. There are number of challenges associated with utilizing cloud computing such as data security, abuse of cloud services, malicious insider, cyber-attacks, and advanced persistent threats (APTs) as discussed in [4, 10, 14, 28, 54, 55]. The various cloud database service providers such as Google Drive, OneDrive, Amazon Cloud Drive, DropBox, etc., [25] are storing the encryption keys in their data centers. According to the 2014 Verizon data breach report [36], databases are one of the most compromised assets. There are some schemes which support certain types of computation on encrypted data for databases. In [3], statistical discloser control techniques are discussed to maintain the privacy of the sensitive data, while allowing the external researcher to compute some statistical values from the patient's databases. Fully homomorphic encryption [6] scheme which supports random calculations on encrypted data. But, this technique is prohibitively slow, therefore, for database intensive applications may not be appropriate. The available proxy-based solutions CryptDB [1], DBMask [7] and non-proxy-based solutions MONOMI [29, 31] which supports to perform computations on encrypted data for relational databases are storing the required keys in the system. If keys are stored, they may create security issues when the server compromises such as cryptanalysis on stored keys, and also issues related to key loss, rollover, backup, recovery, and audit are possible. To mitigate key-related problems, a scheme called 'Stateless Key Management for Proxy-based Systems (SKM-PS)' are presented. In our proposed scheme, the required keys are generated on the fly when user logs in the system, no need to store any key anywhere. Paper is organized into related work, threat model

and security overview of proposed scheme, different possible approaches for protection of third-party managed databases, proposed stateless key management scheme and its performance evaluation, comparison of stateful (stored) and SKM-PS key management, security analysis of SKM-PS scheme, other practical considerations, and conclusion and future scope.

## 2 Related Work

The various key management techniques which protect critical data discussed in literature [18–24] are storing required keys in the system. The key management solutions specially proposed for cloud-based databases in [30, 32–34] are also storing the required encryption/decryption keys in the system. Inspired by CryptDB [1], a proxy-based system, which supports operations on encrypted data at the database server side using SQL aware encryption algorithms with all required encryption keys stored in proxy. Key management of CryptDB [1] is based on chaining of password to the encryption key, which is used for encryption of data for that user. Each user encryption key is protected with his/her own password, and all these keys are stored in the proxy [1]. Using the application login password of user, the stored encryption key of that user is released in the proxy. This concept is known as key chaining. For shared data, CryptDB maintains a SPEAK-FOR relation [1]. CryptDB also uses public-key encryption when user not logged into the system. This system uses both symmetric and public-private key encryption techniques. In this solution, all the keys that are used to apply the security in the system are stored in the proxy.

- In this solution, the keys of unlogged users may be revealed if the proxy is compromised using cryptanalysis techniques by the attacker because keys are stored in the proxy.
- Also this key management scheme suffers from the problems related to key loss, rollover, backup, recovery, and audit.
- The public–private key usage mechanism also adds up overhead in users' transaction used to encrypt and decrypt message key.
- There is an additional overhead of encrypting and storing keys in database. Same amount of overhead is added while retrieving and decrypting the stored encrypted key from database. For example, to store user data in encrypted state, proxy generates a key from user's login credentials, fetches the stored encrypted symmetric key of that user, and decrypts it by generated key. Then, proxy uses user's symmetric key to encrypt the data and sends to database server. So, for executing single transaction, it requires at least three cryptographic operations and two database transactions.
- The various proxy-based security solutions in-line with CryptDB or extension to CryptDB reported in [37–52] are also storing required cryptographic keys in the system.

In [17], voltage security *Inc.* company mentioned about stateless key management solutions but the implementation details are not available. As per Gartner, the proxy-based security solutions (Cloud Access Security Brokers) [35] are going to evolve more and more for the protection of cloud database service. To the best of our knowledge, stateless key management scheme (Not storing keys in the proxy) proposed in this paper is first of its kind for protection of sensitive data using a trusted proxy for cloud-based relational databases. The proposed key management scheme is also applicable to non-relational databases, but the scope of this paper is restricted to relational databases. In most cases, organizations critical data is stored in the relational databases. Therefore, in this paper, our focus is on key management for relational databases.

## *2.1 Threat Model*

The proposed key management scheme addresses the following threats.

i. The first threat is a malicious administrator or IT manager who tries to acquire the stored keys in the proxy by snooping on the key storage server; here, the proposed SKM-PS scheme eliminates the malicious administrator or IT manager from acquiring the stored keys.
ii. The second threat is an external attacker who compromise the proxy. In this case, SKM-PS does not prevent accessing the keys of those users who are already enter into application during an attack, nevertheless ensures the confidentiality of logged-out users keys because keys are not stored in the compromised proxy.
iii. The third threat is related to key loss, key rollover, key backup, key recovery, and audit of the keys; here, the proposed scheme eliminates all these issues by simply not storing the keys. The stateless systems are in general less likely to suffer from the distributed denial of service (DDos) attacks compared with stateful systems.

## *2.2 Security Overview of Proposed Scheme*

In our proposed model, there is a set of users using an application, a trusted proxy, and database server. The database stores encrypted data. Each user's data is encrypted or decrypted with keys generated using login credentials of respective users by trusted proxy. These keys are generated on the fly by trusted proxy when user logs in the system through application and deleted when user logs out from application. In our proposed scheme, since keys are generated dynamically on user login, there is no need to store the keys in the proxy. The issues related to key storage, key loss, key rollover, key backup, key recovery, and audit, thus, becomes null and void with our proposed scheme. There are other types of proxy systems which store keys in encrypted state. But these systems are also susceptible to attacks even though the keys are in encrypted

state. An attacker who gets entry into proxy can try to extract plain keys by employing different cryptanalysis techniques. These cryptanalysis techniques become invalid if the proxy system does not store any key.

## 3 Approaches for Protection of Outsourced Databases

Three approaches are discussed in this section which are used to protect data in the cloud environment.

### *3.1 Approach-1: Security at Client Side to Protect Confidentiality of Data*

As per this scheme, Fig. 1, data is encrypted at the client side before sending it to the cloud database server. All the security modules and keys are with the client. The encrypted data is retrieved from cloud database server and is decrypted at the client side for any manipulations. If there are any modifications in the database, whole database encrypted again and stored. This scheme demands whole database required to be fetched for any OLTP and OLAP transactions, which results in high inefficiency and resource consumption. SPORC [8] and Depot [9] systems follows this scheme.

### *3.2 Approach-2: Security at Server Side for the Protection of Sensitive Data*

In this approach, as shown in Fig. 2, the required security modules resides in the database server system. In this scheme, data decrypted in the server for execution of query and keys is managed at the server side. Unencrypted data is accessible to

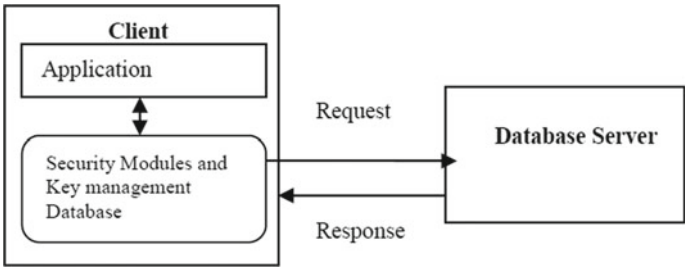



**Fig. 1** Security module at client side

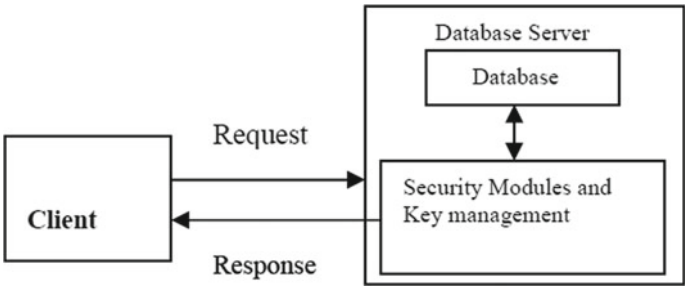



**Fig. 2** Security module at database server side

third-party cloud database service providers during the computations. So, malicious administrator may exploit data. Oracle's TDE [11] and SUNDR [27] follow this scheme.

### *3.3 Approach-3: Security Modules Reside in Proxy*

As per this approach, the client, proxy, and database server are places as shown in Fig. 3. In proxy, data is encrypted before storing into database server. The security mechanisms which allows directly execution of queries on encrypted. This approach does not demand any changes in database server and client applications. The entire cryptographic keys are managed by the proxy which eliminates the need to share or store any keys with database server. This approach promises confidentiality of sensitive data from malicious administrators. Based on this approach, CryptDB has presented the technique to preserve the data confidentiality for SQL database applications. Cipher cloud [12] and Navajo systems [13] follow this scheme. Among the

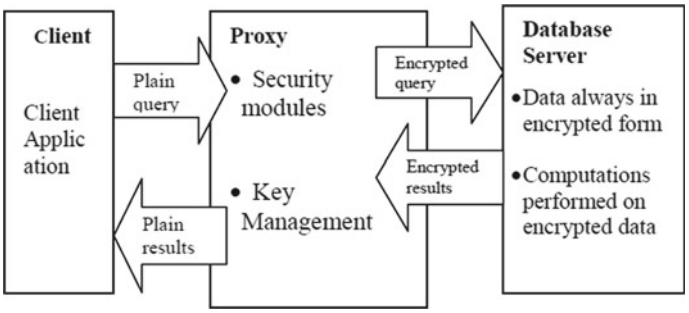



**Fig. 3** Security module at intermediate (proxy) Level

discussed four approaches, the approach four(proxy based) looks very promising for providing the security to the outsourced sensitive data in an untrusted service environment.

In this approach, neither client application nor database schema is required to change. The solution is completely transparent to user and database service providers.

# 4 Proposed Stateless Key Management Scheme for Protection of Outsourced Databases Using TRUSTED Proxy

## *4.1 Architecture Overview*

This para describes the overview of proposed stateless key management as depicted in Fig. 4. The main idea of this scheme is to generate the required keys on the fly when user logs in the system and manage the usage of these keys to protect sensitive data in the proxy without storing the keys. There are three logical entities in the proposed SKM-PS scheme such as client, proxy, and database server. The cryptographic keys are generated and managed in the proxy. On user login from an application using username and password, then an SQL request comes from the application to the proxy to fetch that user credentials from database. The proxy captures this SQL statement, and by using that user credentials, proxy creates cryptographic key for that user. The data encryption and decryption take place using this key in the proxy crypto module engine. The proxy generates and manages all the cryptographic keys that

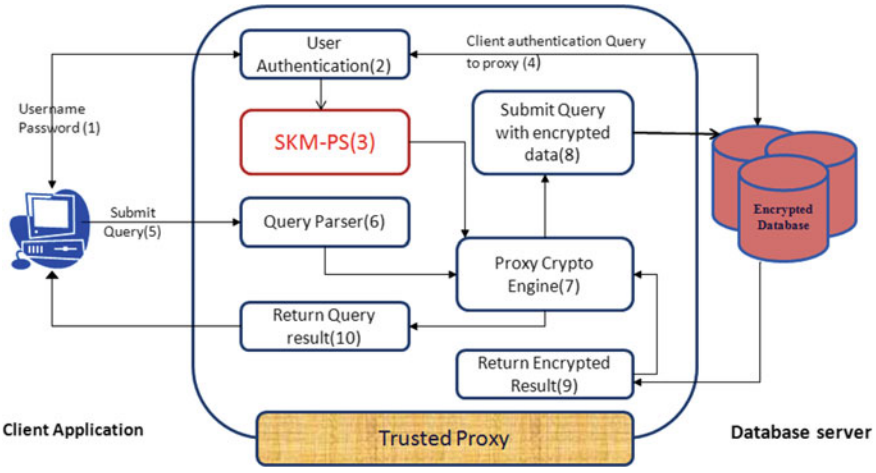



**Fig. 4** Stateless key management scheme for proxy-based systems (SKM-PS) architecture

are used to access data items from the database. The application developer annotates the application's database schema and each user roles and their access rights in the system in the proxy.

## 4.2 Key Derivation Using PBKDF

SKM-PS scheme uses a NIST recommended password-based key derivation function for generating the intermediate in the process of key generation. Key is derived using the user provided password with NIST Special Publication 800-132 [2] as depicted in Fig. 5. The US government uses PBKDF2 to generate strong encryption keys for their systems using user provided passwords. It is a lightweight algorithm that uses only standard and proven hash functions such as NSA's SHA. Key derivation functions are used to derive cryptographic keys information from a secret value, such as a password. Each password-based key derivation function (PBKDF) is defined by the choice of a pseudo random function (PRF) and a fixed iteration count (C). The input to PBKDF includes a password (P), a salt (S), and required length encryption key in bits (kLen).

Symbolically:

$$MK = PBKDF_{(PRF, C)}(P, S, kLen).$$

The minimum length of kLen value is 112 bits. 1000 iteration count is minimum recommended, and for critical keys, iteration count of 100,000 may be appropriate.

## 4.3 Procedure for Generation of Keys in the Proxy

### 4.3.1 Member Key (KM) Generation

i. When user enters credentials in login interface of application, application passes login credentials to proxy in form of SQL statement. The proxy parses the query, extracts the credentials, and generates the required cryptographic key (KM) from credentials using PBKDF2 and AES.

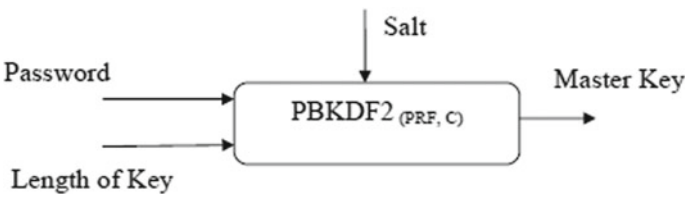



**Fig. 5** Password-based key derivation function 2

$$\left.\begin{array}{ll} \text{K1} & = \text{PBKDF}_{(\text{PRF,C})}(\text{P, S, kLen}) \\ \text{K2} & = \text{truncate}(\text{K1, reqkLen}) \\ \text{KM} & = \text{AES}(\text{K2, keyStr}) \\ \text{P} & = \text{user credentials,} \\ \text{S} & = \text{salt,} \\ \text{kLen} & = \text{key length} \end{array}\right\} \quad (1a)$$

*Truncate function generates the required key of length, reqkLen by truncating K2.*

ii. The generated key (KM) is used for encryption and decryption. It may also be used for obtaining keys for databases, tables, and columns for those user which are logged in the proxy.
iii. Shared sensitive data should be encrypted as per OPEN_TO relation.

### 4.3.2 OPEN_TO: Relation for Shared Data

***OPEN_TO:*** *Indicate that if data 'D' is accessible to the users* $U_1$, $U_2$, … $U_n$, *then data item 'D' is encrypted with a common key (CK) of that user's group.*

▹ **Generation and usage of Common Key (CK) scheme in the proxy for shared data**

This section describes the common key (CK) generation and usage for accessing the shared data in the proxy. The generation of common key (CK) is devised in three different ways. Depending on the user application requirement and availability of the resources, one of the proposed schemes may be adopted to generate the common key in the proxy.

### 4.3.3 Scheme-1: Preparation of Groups and Generation of Their Respective Keys [Member Key and Common Key (CK)]

In the scheme-1, application users in the proxy are categorized into different groups depending on their role. Here, four application user roles are considered in the system for the sake of simplicity. The roles are administrator, manager, accountant, and financier. Administrators can access data owned by managers, accountants, and financiers. Likewise, manager can access accountants' and financiers' data, and lastly, accountant can access financiers' data.

**Definition 4.1** (*Access graph, Fig.* 6) An access graph G having four nodes G1, G2, G3, and G4 each representing a role and $G = (R, D, E)$ having a set of user roles $R$, a set of user's data $D$, and also a set of edges $E$, where an edge $e$ is a pair $(i, j)$ for $i, j \in R$ denoting user $i$ has access to data 'D' owned by user $j$.

Each data item access rights are maintained in the proxy. When a request is received from an application, the proxy first determines the access rights of the user. And accordingly, keys are generated and used in the system for accessing the shared

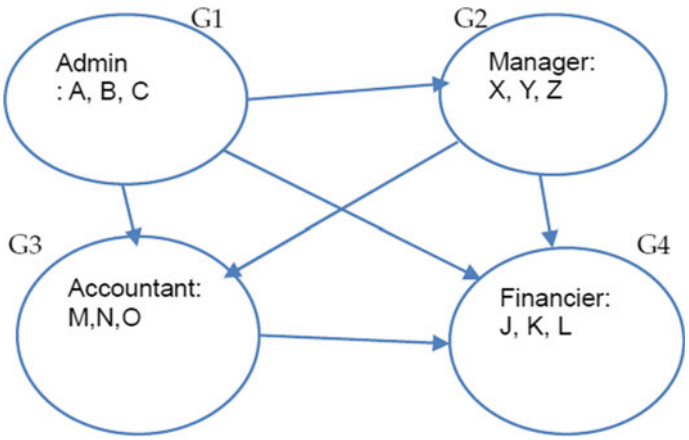



**Fig. 6** Access right graph

data. Table 1 depicts the members of the group, Table 2 depicts the group access rights across other groups, and Table 3 depicts group access rights for common keys.

#### 4.3.4 Preparation of groups and their respective keys (member key and Common key (CK))

i. Initially, administrator annotates one time configuration data to proxy. This data contains information pertaining to users, their roles defining group, relation between groups, and their application login credentials.

**Table 1** Group members

| Group | G1 | G2 | G3 | G4 |
|---|---|---|---|---|
| | A | X | M | J |
| | B | Y | N | K |
| | C | Z | O | L |

**Table 2** Across group access rights

| Group | Access rights | | |
|---|---|---|---|
| G1 | G2 | G3 | G4 |
| G2 | G3 | G4 | – |
| G3 | G4 | – | – |
| G4 | – | – | – |

**Table 3** Group access matrix for common key (CK)

| Group | Common key access rights | | | |
|---|---|---|---|---|
| G1 | CK1 | CK2 | CK3 | CK4 |
| G2 | CK2 | CK3 | CK4 | – |
| G3 | CK3 | CK4 | – | – |
| G4 | CK4 | – | – | – |

ii. Proxy categorizes users into different groups according to the roles defined in the configuration data. And each group is assigned with one random secretID.

$$G_{i=1}^{4} = \text{secretID}_{i=0-4}$$

| Group | secretID |
|---|---|
| G1 | SecretID1 |
| G2 | SecretID2 |
| G3 | SecretID3 |
| G4 | SecretID4 |

iii. Each user has their member key which is generated on the fly by using their application login credentials.
iv. Here, it is to be noted that login credentials are used only one time to generate member keys KM. The assigned secretID$_i$ for each group is stored in the database by encrypting with each group member key and other group member key based on access rights.

***Member key generation***

$$K1 = PBKDF_{(PRF,C)}(P_i, S, kLen)$$
$$K2 = truncate\ (K1, reqkLen)$$
$$KM = AES\ (K2, keyStr)$$

$P_i$ = *user credentials, S = salt, kLen = key length,*
*keyStr = Fixed key string, reqkLen = desired key length*

Truncate function generates the intermediary key having length reqkLen by truncating intermediate key K1. Algorithm AES encrypts fixed key string by the intermediary key K2 to generate member key KM.

**Storing of SecretID in Database**:

$$CiphersecID_i = Enc_{Ki}(secretID_i)$$

*The CipherID$_i$ are stored in the database.*

| Group members | Encrypted secretIDs with member key stored in the database |
|---|---|
| $U_1$ | $CiphersecID_1$ |
| $U_2$ | $CiphersecID_2$ |
| $U_3$ | $CiphersecID_3$ |
| $U_4$ | $CiphersecID_4$ |
| … | … |
| $U_n$ | $CiphersecID_n$ |

v. The proxy generates a common key ($CK_i$) depending on the user belonging group using assigned $secretID_i$ to access shared data in the system. The stored $CipherID_i$ is retrieved from the database and is decrypted using respective user's member key $k_i$ to get corresponding $secretID_i$.
$secretID_i = Dec_{ki}(CipherID_i)$

The common key ($CK_i$) is generated in the proxy in following way

$$\left.\begin{array}{ll} K1 & = PBKDF_{(PRF,C)}(secretID_i, S, kLen) \\ K2 & = truncate(K1, reqkLen) \\ CK_i & = AES(K2, keyStr) \\ S & = salt, \\ kLen & = key\ length, \\ CK & = Common\ Key \end{array}\right\} \quad (2a)$$

vi. The common key ($CK_i$) is used to access the shared data between the users belonging to their group $G_i$.
vii. When user $U_i$ has to access shared data that belongs to other group $G_j$, then common key $CK_j$ belonging to that group is generated by the proxy using that group's $secretID_j$.

In the following example, the encrypted shared data 'xyz' belonging to user Alice from the group manager, i.e., G2. This shared data is OPEN_TO the users John and Dan which belongs to group G1. The data 'xyz' is encrypted with common key of group G2, i.e., CK2. Common key CK2 gets generated by using $secretID_2$ of group G2. This $secretID_2$ is also encrypted with John and Dan member keys and stored in the database. If Alice wants to access the encrypted data item 'xyz,' she logs in the system by using her credentials. The proxy generates Alice member key $KM_3$, retrieves $secretID_2$ of group G2, and decrypts using Alice member key $KM_3$. Using this $secretID_2$, proxy generates the common key ($CK_2$) and access the encrypted data item 'xyz.' If John wanted to read the data item 'xyz,' then proxy generates John member key $KM_4$ using his login credentials and retrieves group G2's $secretID_2$ which is encrypted and stored by the John's member key $KM_4$. Now, proxy decrypts $secretID_2$ and uses to generate the common key ($CK_2$). Using $CK_2$, John decrypts

and reads the data item xyz. The OPEN_TO procedure avoids use of the public-key concept of CryptDB when user is not online. In CryptDB, the shared data 'key' of owner is encrypted and stored with user's public key when she is not online. When user logs in, she decrypts the 'key' using her private key and re-encrypts owner's shared data key with her own symmetric key. When user is not online, two times encryption and decryption of shared data owner's key takes place. In our OPEN_TO relation technique, it is required to decrypt only once to generate CK to access the shared data. Also, OPEN_TO procedure avoids encryption and decryption using asymmetric key. Asymmetric key algorithms are more expensive than symmetric key algorithm.

viii. When user logout from application, all keys will be deleted from the proxy which are belonging to that user.

### 4.3.5 Scheme-2: Member Key and Common Key generation

In scheme-1, there is a 'secretID,' which is used in generation of common key (CK) in database by protecting with member keys which are generated using their login credentials. Scheme-2 avoids the storing of 'secretID' in database. In scheme-2, as depicted in Fig. 7, a 'secretID' for each group is pre-created and shared with all the group members in the system. The user provides secretID using other medium than username and password application interface at the time of login in the system. The secretID along with username is sent to proxy by user via other communication channel such as mobile network. The username and password reach the proxy in the form of SQL statement. Proxy parses the query and creates member key for that user using username and password and also creates common Key (CK) using 'secretID' for accessing shared data. In this scheme, proxy needs to run a daemon which interfaces with other communication channel through which user provides the secretID to the proxy.

i. Member key (KM) is generated using login credentials of user with the help of the procedure explained at (1)

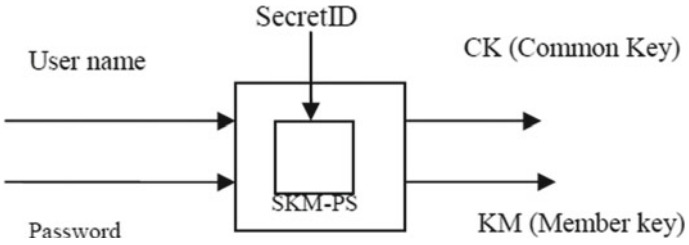



**Fig. 7** Illustration of scheme-2 of SKM-PS

ii. The common key (CK) is generated with the help of the procedure explained at (2).
iii. Shared data can be accessed as per OPEN_TO relation as described in Scheme 1 at points *vi* and *vii*.

All the access rights of the users for shared data are maintained like scheme-1 in the proxy. All the keys are generated on user logs in the system through application. In this scheme, the 'secretID' has to be passed to proxy via other communication channel when user logs in the system along with username and password.

### 4.3.6 Scheme-3: Member Key and Common Key generation

In scheme-3, as depicted in Fig. 8, further the passing of 'secretID' is avoided via other communication channel to proxy for generation of common key (CK)to access the shared data. In this scheme, the password of user is created such that the assigned '$secretID_i$' to the user $U_i$ and user own pass phrase (uP) is concatenated in a fashion to generate password.

Password = '$uP_i$ + $secretID_i$'

i. In this scheme, also users are assigned secretIDs and user's pass phrase (uP) are combined with respective users secretIDs to create a password for that user. When user login the system using this password, the proxy, segregate user's pass phrase, and secretID from the user provided password.
ii. The proxy generates member key (KM) using login credentials (username and pass phrase portion of password (uP) with help of the procedure explained at (1a).
iii. The proxy generates common key (CK) using 'secretID' extracted from user login password with the help of the procedure explained at (2a).
iv. Therefore, this scheme eliminates storing of secretID in the database.

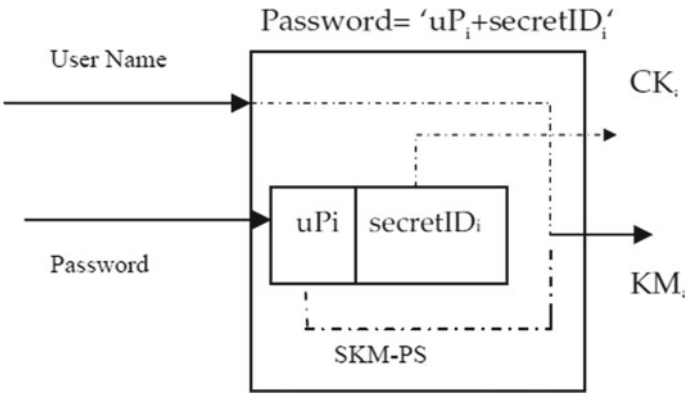



**Fig. 8** Illustration of scheme-3 of SKM-PS

All the users in the proxy as per OPEN_TO relation use CK for encrypting and decrypting shared data. When users logout from application, all the keys including common key are deleted from the proxy. In our proposed key generation scheme, neither user member keys nor common keys are required to store in the proxy. All the required keys are generated on the fly when user login the system. The proposed scheme of key management technique eliminates the requirement of keys storage. Hence, attacks on keys which are stored at proxy level become invalid. These techniques also adhere to many regulatory compliance by not storing the encryption keys. There is no requirement of backup the keys periodically. And there is no fear of keys loosing. A key copying among different servers is not required, and keys can be re-generated or derived in every server as and when it is required dynamically. The proposed stateless key generation procedures is not only applicable to the proxy-based protection system for databases, but also useful in many applications where users are sharing documents.

- **Remarks: Public-Key pair generation using SKM-PS**

  i. The proposed scheme generates the symmetric member key and common key on the user logs in the system. There may be certain application cases where public-key encryption mechanism only applicable to encrypt and decrypt the data in the proxy. In such application scenarios, to generate the member key pair (public and private) for asymmetric algorithms, an alternate syntax is defined; a key pair (KP) is generated using user application login credentials.

$$\textbf{\textit{Member key pair generation}}:$$
$$K1 = PBKDF_{(PRF,C)}(P_i, S, kLen)$$
$$K2 = truncate\ (K1, reqkLen)$$
$$KP = RSA\ (K2, keyStr)$$
$$P_i = user\ credentials,\ S = salt,\ kLen = key\ length,$$
$$keyStr = Fixed\ key\ string,\ reqkLen = desired\ key\ length$$

  ii. To access the shared data, as per OPEN_TO relation, the common key pair (CKP) is generated with PBKDF2 and asymmetric key algorithm using $secretID_i$.

$$\textbf{Common Key Pair Generation}:$$
$$K1 = PBKDF_{(PRF,C)}(\text{secretID}_i,\ S,\ kLen)$$
$$K2 = truncate\ (K1, reqkLen)$$
$$CKP = RSA\ (K2, keyStr)$$
$$S = salt,\ kLen = key\ length,$$
$$CKP = Common\ Key\ Pair$$

For accessing the shared data, secretIDs are generated and assigned to each group as mentioned in scheme-1. These secretIDs are used to generate the common key pair ($CKP_s$) for each group, and $secretID_s$ are encrypted with public key of each group members, and other group members as per data access rights stored in the database. To access the shared data for a user $U_i$ of group $G_i$, the $secretID_i$ is retrieved from the database and decrypted with private key of user $U_i$ and $CKP_i$ is generated using this $secretID_i$, and then shared data is accessed using these keys. Similarly, like scheme-2,

to access the shared data, the secretID is passed using other communication channel to the proxy. Similarly, like scheme-3, the secretID is combined with pass phrase to form a user login password.

### 4.4 Choice of Scheme for OPEN_TO Relation

This paper presented three schemes to generate the common key to access the shared data as per the OPEN_TO relation. The proposed three schemes to generate the common key are relevant in the following three different application scenarios. Therefore, these schemes need to be adopted depending on the application requirements and resource availability.

- **Scenario-1**
  Consider a case where there is no other communication channel (mobile network) is present in the system (client side application) to send secretID to the proxy that is in turn used to generate the common key (CK). In this case, the scheme-1 is relevant where secretID is stored in the database server by encrypting it with the member keys.
- **Scenario-2**
  If application has a provision of using other communication channel (mobile network) to connect the proxy, in this case, the secretID can be obtained from the user via this channel and used to generate the common key (CK).
- **Scenario-3**
  If application has no other communication channel (mobile network) to connect to the proxy and also does not have the provision to store the secretID. Then, the third proposed scheme is useful to generate the common key.

### 4.5 Advantage of SKM-PS Scheme over CryptDB Key Management (stored Key)

Here, it presents advantages of the stateless key management over the stateful key management scheme (CryptDB). The proposed SKM-PS have many advantages over the stateful key management for proxy-based systems. The comparison of these two schemes is shown in Table 4. The summary of the proposed scheme is as follows: It mitigates many issues related to the stored key management of proxy-based systems. On the fly, keys are generated in the proxy at the time of user login into the application. This scheme completely eliminates accessing the keys of unlogged in users in the system by not storing keys in the proxy by the malicious administrators or hackers. Various other key management issues such as key backup, key recovery, key rollover, auditing of the keys, and denial of key access services become null and void with our proposed key management.

**Table 4** Advantage of Stateless Key Management (SKM-PS) over the stored key management (CryptDB)

| Stateful key management (CryptDB) | Stateless key management (SKM-SP) |
|---|---|
| Key is generated at the time of proxy setup and available in the proxy system even after user logout from application | Key is generated when it is required and not available in the proxy after user logout from application |
| Key is required to store | Eliminates key storage |
| Key is used by accessing from the stored location | Key is generated dynamically and used |
| Key is required to backup | Eliminates the key backup |
| There is a possibility of stealing the stored keys using cryptanalysis by malicious administrator or hackers | No cryptanalysis is applicable, since no key is available in the proxy after user logout from application |
| Possibility of insider abuse on stored keys in the proxy | Eliminates insider abuse for the keys which used to encrypt the sensitive data |
| Unlogged user data may be in threat because stored keys may be accessed by the attacker using some cryptanalysis | Eliminates the possibility of accessing the unlogged user data by an attacker |
| Physical attack/damage of keys is possible | Eliminates the physical attack/damage for the keys |
| Does not comply to many security standards | Compliance to many security standards |
| Keys used for encryption of sensitive data may be lost | Guarantees that key cannot be lost |
| Extra infrastructure is required to store the keys and its management | Eliminates extra infrastructure for key storage |
| Key retrieval servers are congenitally in-secure, costly, and highly strain to use | Key recovery is not required; when key is loss, it can be generated using user logs in credentials |
| Key recovery reduces the protection of encryption techniques available to control over decryption data | Not applicable |
| Third-party service may require for the enterprises to manage the keys | No requirement of third-party services for key management |
| Client need authentication to retrieve key from central server | Not applicable |
| Ensuring that keys will still be available for long duration when access to archived data is required will be under question? | Keys can be derived after any number of years having only user login credentials |
| It is very difficult to ensure that an authorized staffer be able to access keys in a disaster when servers must be rebuilt from encrypted backups without the original backup software or tape drive that did the encryption? | Only authorized user with right login credentials can only access to encrypted backup |
| Key recovery systems are particularly vulnerable to compromise by authorized individuals who abuse or misuse their positions | Not possible |
| Performance overhead is higher when accessing the stored key from remote location | Performance of generation of key dynamically is better than accessing and using the stored key |
| Requires dedicated manpower to manage the key backups | Eliminates the need for dedicated IT head count for key management because it removes the need to constantly backup key stores |
| Auditing of the keys required regularly | Eliminates the key auditing |
| There is possibility of denial of service while accessing key from stored location | Avoids denial of service |

### 4.6 Stateless Key Management Performance Statistics

Here, the performance analysis of the stateless key management scheme has been presented. C language is used to implement the presented key management scheme using OpenSSL and Mysql libraries. The experiments are carried on Linux platform running on 16-core Intel Xeon processor @2.60 GHz system with 64 GB of RAM. The key length 128, 256, and 512 bits have been considered to generate. The member key (KM) generation performance timings are shown in Table 5, and common key (CK) generation performance timings are shown in Table 6. The keys are generated when user logs in the system and key remains in the proxy complete session until user logout from the application. Therefore, key generation timings are shown in Tables 5 and 6 realistic and even performing better compared to storing and accessing the key in the proxy. The key generation parameters such as user login credentials are fixed to 8 bytes and secretID for Common key generation is fixed 60 bytes. Table 7 shows accessing and decryption timings of stored encrypted key in the proxy. The member key and common key are generated using AES algorithm along with PBKDF2. The default parameters of PBKDF2 such as number of rounds 1000, 128 bits salt, and SHA512 are used. In this implementation, all three entities such as application, trusted proxy, and database are residing in the local system. Apart from avoiding the key-related issues with stored key in the proxy, the proposed SKM-PS scheme is performing better timings in key generation than stored key access process. The performance comparison of SKM-PS and stored key usage in the proxy are depicted in Tables 8 and 9, and Fig. 13 shows graphical representation. Figures 9, 10, and 11 depict the timings of member key, common key, and decryption of stored key, respectively (Fig. 12).

**Observations**

→ Performance overhead is higher in stored key management (CryptDB) compared to stateless key while data encryption.

→ Apart from mitigating the various key management issues with stored key, the

**Table 5** Performance statistics of member key generation using SKM-PS

| S. No. | Key length | Key generation timing (s) |
|---|---|---|
| 1 | 128 | 0.004457 |
| 2 | 256 | 0.008911 |
| 3 | 512 | 0.017732 |

**Table 6** Performance statistics of common key (CK) generation timings using SKM-PS

| S. No. | Key length | Key generation timing (s) |
|---|---|---|
| 1 | 128 | 0.004885 |
| 2 | 256 | 0.009312 |
| 3 | 512 | 0.018093 |

**Table 7** Performance statistics of decryption of stored key in the proxy

| S. No. | Key length | Key generation timing (s) |
|---|---|---|
| 1 | 128 | 0.019698 |
| 2 | 256 | 0.021047 |
| 3 | 512 | 0.024605 |

**Table 8** Performance comparison of stored key and member key(SKM-PS)

| Key length | Stored key decryption in the proxy | SKM-PS member key generation |
|---|---|---|
| 128 | 0.019698 | 0.004457 |
| 256 | 0.021047 | 0.008911 |
| 512 | 0.024605 | 0.017732 |

**Table 9** Performance comparison of stored key and common key of SKM-PS

| Key length | Stored key decryption in the proxy | SKM-PS Common key generation |
|---|---|---|
| 128 | 0.019698 | 0.004885 |
| 256 | 0.021047 | 0.009312 |
| 512 | 0.024605 | 0.018093 |

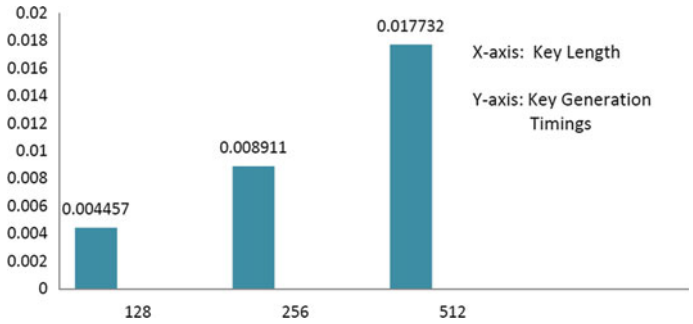



**Fig. 9** Member key (KM) generation of 128, 256, and 512 bits timings in seconds

proposed stateless key management achieves better performance gain compared to stored key management. The performance gain is shown in Table 10, and same has been illustrated in Fig. 14. The performance gain is higher side with lower key length and lower in higher key length.

→ The number of cryptographic operations and database transaction required to store a given user data item (for example) in an encrypted state with CryptDB key management, and proposed stateless key management scheme is discussed below.

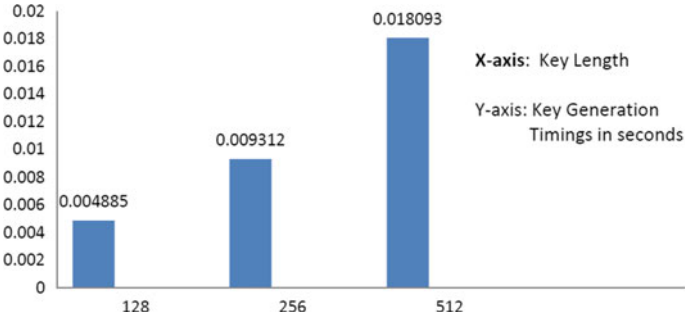



**Fig. 10** Common key (KM) of 128, 256, and 512 bits generation timings in seconds

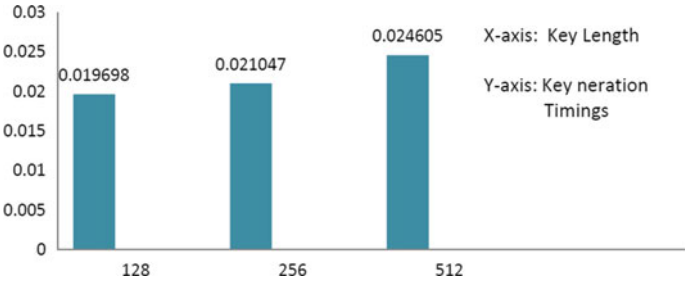
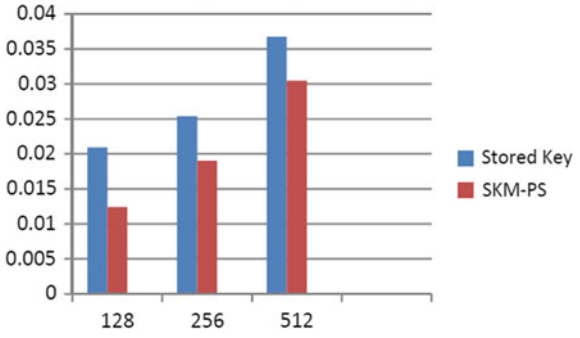



**Fig. 11** Decryption timings of stored key in seconds

**Fig. 12** Performance Comparison of common key and stored key. *X*-axis: key length and *Y*-axis: timings in seconds




## With CryptDB

- Proxy generates a key from user's login credentials.
- Fetches the stored encrypted symmetric key of that user and decrypts it using generated key.
- Then proxy encrypts data using user's symmetric key and dispatches to database server to store.

**Table 10** Performance gain with SKM-PS compared to stored key

| S. No. | Key length | % of gain with SKM-PS |
|---|---|---|
| 1 | 128 | 441.9 |
| 2 | 256 | 236.19 |
| 3 | 512 | 138.76 |

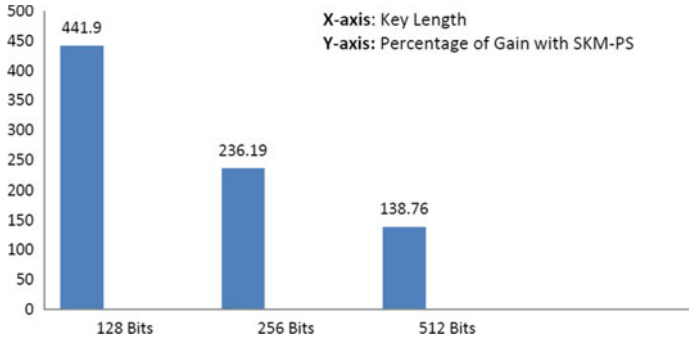



**Fig. 13** Performance gain with SKM-PS compared to stored key for keys length 128, 256, 512 bits

- So, for executing single transaction, it requires at least three cryptographic operations and two database transactions.

**With SKM-PS**

- Proxy generates key from user provided login credentials.
- User data item is encrypted using this key, and proxy sends encrypted data to the database server to store.
- Therefore, SKM-PS requires only two cryptographic operations and one database transaction.

The number of cryptographic operations and database transactions required with respect to CryptDB and SKM-PS is shown in Table 11.

**Table 11** Number of cryptographic and database transactions requirement for encryption of a data item comparison

| Scheme | No. of cryptography ops | No. of database transactions |
|---|---|---|
| Stored key (CryptDB) | 3 | 2 |
| SKM-PS | 2 | 1 |

## 5 Security Analysis of Proposed Stateless Key Management

The proposed stateless key management scheme is provable secure for generating the cryptographic keys and their use for preserving the database confidentiality in an untrusted service environment for SQL databases make use of the proxy methodology. Intuitively, security property from the SKM-PS scheme expected is a secure key generation process; i.e., keys generated using the proposed scheme is random keys, and it is not possible to guess or derive keys by any adversary in the proxy for those users who are not logged into the system.

**Definition 5.1** (*Secure Key Generation*) The key generation scheme 'KG' is secure such that an adversary '*ADV*' does not able to generate or derive the cryptographic keys 'K' in the proxy without knowing the login credentials of the authentic users in the proxy.

***Proof*** The key generation scheme 'KG' is $(q, t, \varepsilon)$-secure, if for all $t$-time adversaries A that send at most q queries to the key generation Oracle $O(P)$ it holds that $\Pr[A^{O(P)} \rightarrow K] \leq \varepsilon$.

The probability of an adversary generating/deriving cryptographic key in the proxy using the random login credentials is almost negligible.

The proposed SKM-PS scheme uses PBKDF2 and AES-512 cryptographic algorithms to generate member key and common key in the proxy to access owned and shared data, respectively. The PBKDF2 is used to generate the intermediate key say $K_1$, and the final member and common key are generated using the AES block cipher. The PBKDF2 uses SHA512 hash function, it is a 'one way' function meaning that it is very hard to 'invert' the function, i.e., to find the original message given only a message digest. It is hard to find two different messages $(M, M')$ such that Hash($M$) = Hash($M'$).

I. **Hash function**:

A hash function is a deterministic function which maps a bit string of an arbitrary length to a hashed value which is a bit string of a fixed length. Let h denote a hash function whose fixed output length is denoted by $|h|$. This h is one way function.

**Definition 5.2** A function $h : X \rightarrow Y$ is one way if $h(x)$ can be computed efficiently for all $x \in X$, but $f^{-1}(y)$ cannot be computed efficiently for ${}_{y \in R} Y$.

In this definition, $X$ represents the domain of the function $h$, $Y$ represents the range, and the expression ${}_{y \in R} Y$ stands for a $y$ that is randomly chosen from $Y$. Consequently, it must be efficiently compute $h(x)$ for all $x \in X$, whereas it must not or only with a negligible probability be possible to compute $f^{-1}(y)$ for randomly chosen $y \in Y$.

II. **Pseudo Random Permutation (AES Block cipher)**:

**Definition (Pseudo Random Permutation) 5.3**: A family of permutations $\{f_s : \{0, 1\}^n \rightarrow \{0, 1\}^n\}$ is a strong PRP family (also known as block cipher) if it is:

A. **Efficiently computable**: There is a deterministic polynomial-time algorithm $F$ such that F(s,x)=$f_s(x)$ for all choices of seed s and input $x \in \{0, 1\}^n$.

B. **Pseudo-random**: For every non-uniform probability polynomial time(n.u.p.p.t) A $\mid \Pr_{f \leftarrow \{f_s\}}[A^{f,f^{-1}}(1^n) = 1] - Pr_{F \leftarrow P(\{0,1\}^n}[A^{F,F^{-1}}(1^n) = 1] \mid = \text{negl}(n)$. Here, $P(\{0, 1\}^n)$ is the set of all permutations from $\{0, 1\}^n$ to $\{0, 1\}^n$, and $A^{f,f^{-1}}$ denotes that the algorithm A has oracle access to both the function f and its inverse.

**Symmetric Key security definition 5.4**: A symmetric encryption scheme (S) with message space M and a cipher text space C is perfectly secrete if for all $m_0, m_1 \in$ M and all $c \in C$, $\Pr[\text{Enc}_k(m_0) = c] = \Pr[\text{Enc}_k(m_1) = c]$. That is, the distributions of $\text{Enc}_k(m_0)$ and $\text{Enc}_k(m_1)$ are identical, over the choice of the key $k$.

Let us say, an adversary '*ADV*' has compromised the proxy using some trapdoor in it. Now question arise that 'Can *ADV* able to generate or derive the cryptographic key to gain access to the data belonging to application's users?' This question can be answered with two cases. In first case, there is no live users, i.e., no user is logged in application; in this case, proxy does not have any credentials nor it has any users' member key or group key. So, *ADV* will not be able to generate any key as proxy does not have user credentials, and also *ADV* will not be in state of guessing the credentials by any cryptanalysis techniques, since proxy does not store any keys. While in second case, if there are users logged in, *ADV* may have access to keys for those users as long as they are active and not logged out. But again *ADV* does not have access to the keys or credentials of unlogged users as proxy destroys the keys as and when user logs out of the application. The generation of right cryptographic keys depends on right login credentials of the application users in the proxy. Any adversary is unable to generate the keys in the proxy without having the correct login credentials for the application users. The proof of this concept is shown using random oracle model.

**Random Oracle Model**

A random oracle [53] is an ideal primitive which provides a random output for each new query. Identical input queries are given the same answer. A random permutation is an ideal primitive that contains a random permutation $P : \{0, 1\}^n \rightarrow \{0, 1\}^n$. The ideal primitive provides oracle access to $P$ and $P^{-1}$. An ideal cipher is an ideal primitive that models a random block cipher $E : \{0, 1\}^k x\{0, 1\}^n \rightarrow \{0, 1\}^n$. Each key $k \in \{0, 1\}^k$ defines a random permutation $E_k = E(k, .)$ on $\{0, 1\}^n$. The ideal primitive provides oracle access to $E$ and $E^{-1}$; that is, on query $(0, k, m)$, the primitive answers $c = E_k(m)$, and on query $(1, k, c)$, the primitive answers m such that $c = E_k(m)$. These oracles are available for any $n$ and $k$. The function $f$ is deterministic, so the bit string $f(s_0)$ dependent only on the seed. The generated bit strings from $f(s_0)$ should look like truly random bits, given the seed is chosen at random. If this property holds, then the bit generator is secure.

The two algorithms (PBKDF2 and AES) which are used in the scheme are random oracles, therefore, predicting output of these two algorithms probability is negligible. These algorithms (AES and PBKDF2) produce a unique random output value from a unique input (login credentials) and changes randomly when there is any change in the input (login credentials) and are deterministic algorithms. In addition, these

two algorithms are NIST standard security algorithms, and their security proof is described above. Therefore, any adversary 'ADV' in the proxy cannot able to generate the authentic cryptographic keys to access the sensitive data in the system without having right login credentials. Hence, the key generation scheme proposed in the section is secure as long as long as input parameters such as username, password, and secretID are maintained confidentially.

## 6 Other Practical Considerations

### *6.1 Number of Rounds Selection for PBKDF2*

The maximum number of rounds needed to opt depends upon the system capability and performance required by application. In PBKDF2, as number of rounds increases performance decreases. Let us consider: The time is required to generate an intermediate key K1 using user login credentials is 'tv' on a normal system where the proxy program is running. This 'tv' can be adjusted with the number of rounds in PBKDF2. Let say, there is a threat scenario where following assumptions are made:

i. A potential attacker has processing power 'ft' times higher CPU speed than the system where proxy program is running.
ii. A user password has n-bits of entropy, which means that an attacker may try to deduct a user password with dictionary of plausible passwords that takes on average of $2^{n-1}$ tries.
iii. An attacker may try to crack the system to find the password with time less than 'p' where p denotes the attacker's patience.

Therefore, the time required to break a single password is set to exceed the attacker's patience 'p'.

$$\mathrm{tv} \cdots 2^{n-1} > \mathrm{ft} \cdots \mathrm{p}$$

As password entropy bits '$n$' increases, it becomes harder for the attacker to crack the password. Hence, it is better to use higher number of rounds to protect password guessing from dictionary attacks.

### *6.2 Password Change*

Change of password only affects the member key (KM) in our proposed scheme. The new member key has to be generated for that user using new password, and any data encrypted with member key generated with old password has to be decrypted and re-encrypted with newly generated member key. Also, secretID encrypted with old member key has to be decrypted and re-encrypted with newly generated member key.

### 6.3 Forget Password

Passwords and secretIDs used for generation of keys need to securely protected. A password forget mechanism need to setup.

### 6.4 Adding New Application User

To add new user, the proxy should know OPEN_TO relation with other users for accessing of the common/shared data in the system. Basically, the proxy should know that new user belongs to which group in the system. Accordingly, the proxy maintains the access rights in the system.

### 6.5 Proxy Deployment Scenarios

The proxy can be deployed (shown in Fig. 14) within the organization premises and also in the cloud. If proxy is setup on the cloud, organization's administrator should manage it not the service provider. Required security settings for sensitive data need to be configured in the proxy by the proxy administrator.

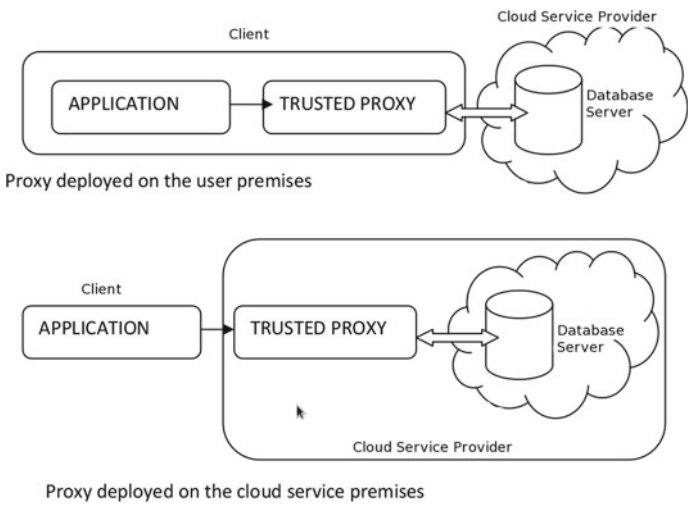



**Fig. 14** Proxy deployments

## 7 Conclusions and Future Scope

Key management has become a crucial security concerns while using the third party services like public cloud databases. Proxy-based security solutions are more prominent way of protecting the confidentiality of sensitive data in databases while in use. In previous proxy-based solutions, which protects confidentiality of sensitive data for relational databases by performing operation directly data which is encrypted in the database, required cryptographic keys to be stored in the proxy. The adversaries may steal the keys by compromising the proxy with help of various cryptanalysis techniques. The adversaries may retrieve the keys of unlogged users. The various issues which are related to key storage, key lost, key rollback, etc., may also arise. A stateless key management scheme is proposed for proxy-based security solutions (SKM-PS) where required cryptographic keys are generated on the fly in the trusted proxy. The scheme presented in this paper generates the required cryptographic keys to be used in the proxy using login credentials along with 'secretID' supplied by user from application dynamically (on the fly). These keys are used to encrypt and decrypt the sensitive data in the proxy. Our proposed key generation scheme eliminates storing of keys in the proxy, hence, avoids cryptanalysis on the stored keys in the proxy (attacks on cipher text) and other key-related issues such as key loss, rollover, backup, recovery, and audit. Required keys are generated on the fly no need to store any keys in the proxy, and all the keys will be deleted in the proxy once user logout from application which is contrary to the cryptDB. Our proposed solution also satisfies many regulatory compliance by not storing the keys anywhere. As a future plan, the possibility of extending the stateless key management scheme in distributed proxy-based protection systems for databases can be explored.

## References

1. Popa RA, Redfield CMS, Zeldovich N, Balakrishnan H (2011) CryptDB: protecting confidentiality with encrypted query processing. In: SOSP 11, Cascais, Portugal, 23–26 Oct 2011
2. Turan MS, Barker E, Burr W, Chen L (2010) NIST Special Publication 800-132: Recommendation for Password-Based Key Derivation. Part 1: Storage Applications
3. Herranz J, Nin J (2014) Secure and efficient anonymization of distributed confidential databases. Int J Inf Secur 13(6):497–512
4. Gonzales D, Kaplan J, Saltzman E, Winkelman Z, Woods D (2015) Cloud-Trust—a security assessment model for infrastructure as a service (IaaS) clouds. IEEE Trans Cloud Comput PP(99):1-1. https://doi.org/10.1109/TCC.2015.2415794
5. http://townsendsecurity.com/products/encryption-key-management
6. Gentry C (2009) Fully homomorphic encryption using ideal lattices. In: STOC, pp 169–178
7. Sarfraz MI, Nabeel M, Cao J, Bertino E (2015) DBMask. In: Proceedings of CODASPY15, San Antonio, TX, USA, 02–04 March 2015
8. Feldman AJ, Zeller WP, Freedman MJ, Felten EW (2010) SPORC: group collaboration using untrusted cloud resources. Princeton University
9. Mahajan P, Setty S, Lee S, Clement A, Alvisi L, Dahlin M, Walfish M (2011) Depot: cloud storage with minimal trust. The University of Texas at Austin

10. Younis YA, Kifayat K, Merabti M (2014) An access control model for cloud computing. J Inf Secur Appl 19(1):45–60
11. An Oracle White Paper (2012) Oracle advanced security transparent data encryption best practices
12. http://www.ciphercloud.com/technologies/encryption/
13. https://securosis.com/tag/navajosystems
14. Buyya R (2013) Introduction to the IEEE Transactions on Cloud Computing. IEEE Trans Cloud Comput 1(1)
15. http://www.verizonenterprise.com/pcireport/2014/
16. http://web.townsendsecurity.com/encryption-key-management-resources/
17. http://www.voltage.com/technology/stateless-key-management/
18. Seitz L, Pierson J, Brunie L (2003) Key management for encrypted data storage in distributed systems. In: IEEE Proceedings of the second security in storage workshop (SISW), 2003
19. Hur J, Yoon H (2010) A multi-service group key management scheme for stateless receivers in wireless mesh networks. J Mob Netw Appl 15(5):680–692
20. Liu D, Ning P, Sun K (2003) Efficient self-healing group key distribution with revocation capability. In: Proceedings of the 10th ACM conference on computer and communications security, Washington, DC, USA, 27–30
21. Hur J, Shin Y, Yoon H (2007) Decentralized group key management for dynamic networks using proxy cryptography. In: Proceedings of the 3rd ACM workshop on QoS and security for wireless and mobile networks, Chania, Crete Island, Greece, 22–24, 2007
22. Catuogno L, Galdi C (2014) Analysis of a two-factor graphical password scheme. Int J Inf Secur 13(5):421–437
23. Rehana Y, Eike R, Guilin W (2014) Provable security of a pairing-free one-pass authenticated key establishment protocol for wireless sensor networks. Int J Inf Secur 13(5):453–465
24. Eschenauer L, Gligor VD (2002) A key-management scheme for distributed sensor networks. In: CCS02 proceedings of the 9th ACM conference on computer and communication security, New York, USA, 2002
25. www.cnet.com/how-to/onedrive-dropbox-google-drive-and-box-which-cloud-storage-service-is-right-for-you/
26. Security in cloud computing (2014) Int J Inf Secur 13(2):95–96
27. Li J, Krohn M, Mazieres D, Shasha D (2004) Secure untrusted data repository (SUNDR). NYU Department of Computer Science
28. Qian H, Li J, Zhang Y, Han J (2015) Privacy-preserving personal health record using multi-authority attribute-based encryption with revocation. Int J Inf Secur 14(6):487–497
29. Tu S, Kaashoek MF, Madden S, Zeldovich N (2013) Processing analytical queries over encrypted data. In: PVLDB 2013. VLDB Endowment, pp 289–300
30. Damiani E, De Capitani di Vimercati S, Foresti S, Jajodia S, Paraboschi S, Samarati P (2005) Key management for multi-user encrypted databases. In: Proceeding StorageSS '05 proceedings of the 2005 ACM workshop on storage security and survivability
31. Ferretti L, Pierazzi F, Colajanni M, Marchetti M (2014) Scalable architecture for multi-user encrypted SQL operations on cloud database services. IEEE Trans Cloud Comput 2(4):448–458
32. Sun X-h Advances in technology and management. Advances in intelligent and soft computing, vol 165, pp 315–319
33. Lanovenko A, Guo H (2007) Dynamic group key management in outsourced databases. In: Proceedings of the world congress on engineering and computer science 2007 WCECS 2007, San Francisco, USA, 24–26, 2007
34. Bennani N, Damiani E, Cimato S (2010) Toward cloud-based key management for outsourced databases. In: Computer software and applications conference workshops (COMPSACW), 2010 IEEE 34th annual, 19–23, 2010, pp 232–236
35. http://www.gartner.com/it-glossary/cloud-access-security-brokers-casbs/
36. http://www.imperva.com/docs/wp_topten_database_threats.pdf
37. Google. Encrypted big query client. https://code.google.com/p/encrypted-bigquery-client/

38. Patrick G, Martin H, Isabelle H, Florian K, Mathias K, Andreas S, Axel S, Walter T (2014) Experiences and observations on the industrial implementation of a system to search over outsourced encrypted data. In: Lecture notes in informatics. Sicherheit
39. Kepner J, Gadepally V, Michaleas P, Schear N, Varia M, Yerukhimovich A, Cunningham RK (2014) Computing on masked data: a high performance method for improving big data veracity. CoRR
40. sql.mit.edu is a SQL server at MIT hosting many MIT-ran applications
41. Arasu A, Blanas S, Eguro K, Kaushik R, Kossmann D, Ramamurthy R, Venkatesan R (2013) Orthogonal security with Cipherbase. In: Proceedings of the 6th biennial conference on innovative data systems research (CIDR), 2013
42. Arasu A, Eguro K, Kaushik R, Kossmann D, Ramamurthy R, Venkatesan R (2013) A secure coprocessor for database applications. In: Proceedings of the 23rd international conference on field programmable logic and applications (FPL), 2013
43. Tu S, FransKaashoek M, Madden S, Zeldovich N (2013) Processing analytical queries over encrypted data. In: International conference on very large databases (VLDB), 2013.151
44. Tetali SD, Lesani M, Majumdar R, Millstein T (2013) MrCrypt: static analysis for secure cloud computations. In: Proceedings of the 2013 ACM SIGPLAN international conference on object oriented programming systems languages & applications (OOPSLA), 2013
45. Kepner J, Gadepally V, Michaleas P, Schear N, Varia M, Yerukhimovich A, Cunningham RK (2014) Computing on masked data a high performance method for improving big data veracity. CoRR
46. Stephen JJ Savvides S, Seidel R, Eugster P (2014) Practical confidentiality preserving big data analysis. In: HotCloud
47. Tople S, Shinde S, Chen Z, Saxena P (2013) AUTOCRYPT: enabling homomorphic computation on servers to protect sensitive web content. In: Proceedings of the 20th ACM conference on computer and communications security (CCS), 2013
48. Ayday E, Raisaro JL, Hengartner U, Molyneaux A, Hubaux J-P (2013) Privacy-preserving processing of raw genomic data. EPFL-REPORT-187573
49. Nabi Z, Alvi A (2014) Clome: the practical implications of a cloud-based smarthome. CoRR, abs/1405.0047
50. Corena JC, Ohtsuki T (2012) Secure and fast aggregation of financial data in cloud-based expense tracking applications. J Netw Syst Manag
51. Hummen R, Henze M, Catrein D, Wehrle K (2012) A cloud design for user-controlled storage and processing of sensor data. In: Proceedings of the 4th IEEE international conference on cloud computing technology and science (Cloud Com), 2012
52. Kagadis GC, Kloukinas C, Moore K, Philbin J, Papadimitroulas P, Alexakos C, Nagy PG, Visvikis D, Hendee WR (2013) Cloud computing in medical imaging. In: Cloud computing in medical imaging
53. Bellare M, Rogaway P (1993) Random oracles are practical: a paradigm for designing efficient protocols. In: Proceedings of the 1st ACM conference on computer and communications security, pp 62–73
54. Subarana S (2019) An efficient security framework for data migration in a cloud computing environment. J Artif Intell 1(01):45–53
55. Adithya M, Scholar PG, Shanthini B (2020) Security analysis and preserving block-level data DE-duplication in cloud storage services. J Trends Comput Sci Smart Technol (TCSST) 2(02):120–126

# Exploration of Blockchain Architecture, Applications, and Integrating Challenges


**Jigar Mehta, Nikunj Ladvaiya, and Vidhi Pandya**

**Abstract** Internet is the technology that is highly used for communication as well as data transmission in the current scenario. Because of the rapid growth in such technology, the issues related to security are also increasing. Blockchain is emerging as a revolutionizing technology in this field. Basically, blockchain relies on the four components, namely decentralization, anonymity, persistence, and audibility. It focuses more on protecting data from unauthorized parties in performing any modification to the existing data. Blockchain technology is also defined as a digital ledger. It is not only limited to cryptocurrencies but with the potential to be transparent and fair, which opens many doors to various technologies like IoT, big data, and many more. This paper includes blockchain locution, and it will encounter some typical consensus algorithms that are ought to analyze blockchain applications and some technical challenges. The proposed research work will also concentrate more on the recent advances to tackle those challenges.

**Keywords** Blockchain · IoT · Big data · Consensus algorithms · The architecture of blockchain · Challenges in blockchain · Applications of blockchain

## 1 Introduction

Day by day, the use of the Internet increases very rapidly; while observing the emerging fields like IoT and big data, it can be said that a massive amount of data is produced every day; wherever data is available, it initiates the need to protect by applying various security measures. These security algorithms will get complicated day by day, and it becomes even harder to implement. This is where blockchain comes into the picture, where it removes any non-trustworthy third party that will solve 2

J. Mehta (✉) · N. Ladvaiya
Department of Information Technology, Devang Patel Institute of Advance Technology and Research, CHARUSAT, Changa, Gujarat, India

V. Pandya
Department of Computer Science & Engineering, Devang Patel Institute of Advance Technology and Research, CHARUSAT, Changa, Gujarat, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_44

purposes: The implementation of the security algorithms is easy, and the transaction takes place directly between the sender and receiver to reduce the complexity. The basic goals or pillars of security are confidentiality, integrity, and availability. Confidentiality means none other than the verified and receiving authority can have access to the information. Integrity means no third party should have the right to change the data, and the last term availability means the information should be available to the verified person present anywhere and anytime. From these three pillars, blockchain focuses more on integrity by maintaining the chain of blocks, where each block contains the hash value of the past block, and thus it creates the chain.

Blockchain was first proposed in 2008 and implemented in 2009 [1]. It is generally considered as a secure digital ledger. It was first recognized by one of its applications that is Bitcoin created by Satoshi Nakamoto by making Bitcoin as one of the first applications of blockchain [2]. Bitcoin was first created to remove a third party such as a bank from a transaction and create a peer-to-peer transaction system. The reason for removing the third party is to solve an age-old human issue called trust. Due to the huge popularity of Bitcoin, more people were interested in blockchain and a vast number of its applications were discovered.

As mentioned earlier, blockchain stores information of all the transactions in the form of a group of blocks. Therefore, the working of blockchain is as follows: The sender broadcasts the new transaction into the network. But yet the transaction is new and is not confirmed. Once the nodes receive the transaction, they will validate it will store a copy of it in their transactional pool. After validation, some special nodes called miners will create a block and then this block goes through several operations which are known as mining. The first block to complete the mining will be added to the chain, thus forming a chain of transactions or blocks known as blockchain [2].

This paper is more focused on the challenges faced in various applications generated from blockchain after its integration with several other fields of technology, namely the Internet of things and big data. Further, this research work will analyze different ways of integration and study the challenges and its potential benefits. The manuscript is separated as follows: Sect. 2 describes blockchain and its architecture, Sect. 3 explains distinctive consensus algorithms, Sect. 4 illustrates different uses of blockchain, Sect. 5 defines difficulties in integration with various technologies and possible solutions, and Sect. 6 gives the conclusion.

## 2 Blockchain: Overview and Architecture

In 2008, Satoshi Nakamoto distributed a paper containing an inventive electric shared money framework known as Bitcoin by eliminating the outsider along these lines permitting just the two consenting partakers to execute straightforwardly [3]. Bitcoin demonstrated an appropriate case of hashing calculations and unbalanced cryptosystems cooperating. Here, a huge public ledger is created by the volunteer nodes and this public lecture is responsible for tracking every transaction taking place in the network. This public ledger is known as a blockchain. Our typical blockchain algorithms can

**Table 1** Comparison between types of blockchain based on participation of no. of nodes

| | Public | | Consortium | | Private | |
|---|---|---|---|---|---|---|
| | All nodes | Selected nodes | All nodes | Selected nodes | All nodes | Selected nodes |
| Consensus determination | √ | | | √ | | √ |
| Read permission | √ | | | √ | | √ |

**Table 2** Comparison between types of blockchain based on performance

| | Public | | Consortium | | Private | |
|---|---|---|---|---|---|---|
| | Low | High | Low | High | Low | High |
| Integrity | | √ | √ | | √ | |
| Efficiency | √ | | | √ | | √ |

be categorized into 3 parts, i.e., public blockchain, private blockchain, and consortium blockchain. In public blockchain, all the nodes can take part in the consensus process. In private blockchain, only the nodes from the specific controlling organization are allowed to take part in the consensus process. In consortium blockchain, only some preselected nodes are allowed to take part in the consensus algorithm. Tables 1 and 2 represent a detailed comparison of different types of blockchain.

Blockchain technology is a combination of 6 key features [4] which is explained below.

**Decentralized**: As discussed earlier, blockchain is said to be decentralized because it does not have any central node or a governing node. Each node has a copy of the chain of transactions.
**Unambiguous**: All the data blocks are transparent to all the other nodes.
**Open Source**: Another bit of leeway of blockchain is that it is open source; hence, anybody can utilize blockchain and make their application.
**Autonomy**: Blockchain can be said autonomy because every node that participates in the chain can transfer or update data securely.
**Unchangeable**: When a transaction takes place, its records are stored permanently into the chain, and it cannot be changed unless any node takes 51% of the ownership.
**Anonymity**: As blockchain supports peer-to-peer transaction system, it can also perform anonymous transactions where the information known is the person's blockchain address [4].

Here, some acronyms are defined for convenience:

Ver: Version
PBH: Previous block hash
TS: Timestamp

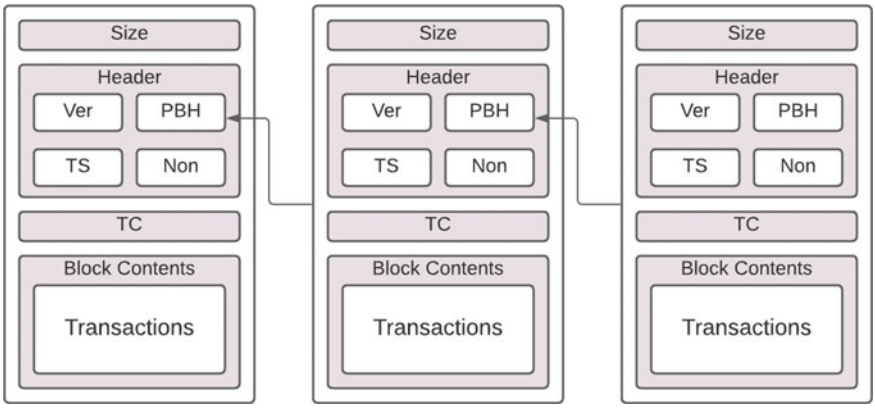



**Fig. 1** An illustration of blocks in blockchain architecture

Non: Nonce
TC: Transaction counter.

An architecture of blockchain is shown in Fig. 1. As mentioned earlier that blockchain is a combination of several blocks which is represented for $N$ number of blocks in Fig. 1. Each block contains the complete information of transactions taking place between two entities. You can see from the figure that a typical blockchain block is divided into four subparts, i.e., size, header, transaction counter, and block contents, where the first part constitutes of size of the block, in bytes. Then, the header part constitutes 4 different parts, i.e., version, hash value, timestamp, and nonce. The version represents the version of the protocol; then, the hash value represents the hash of the past block in the chain. For example, as shown in the figure block $N + 1$ has stored the hash value of block $N$. The timestamp includes the time when the block was created and the nonce is the counter used while calculating the proof of work in the algorithm. The transaction counter gives information regarding the number of transactions that are stored in that particular block. And at last, all the contents of the transactions are stored including sender information, receiver's information, transaction amount/items, etc. Since each node has a copy of the blockchain, there is a possibility of some malicious activity by some of the nodes. To avoid that, some algorithms were introduced such as proof of work (PoW) and proof of stake (PoS) which will be discussed in Sect. 3.

## 3 Blockchain: Consensus Algorithms

In blockchain, the cycle to arrive at an agreement between deceitful nodes is a change of the Byzantine general issue [4]. In Byzantine problem, a gathering of officers who order a specific aspect of the military circle the city to be captured. But

every general has to attack at the same time or they will not be able to win the war. So how would they reach a consensus of whether to attack or to retreat? Therefore, some of the approaches are mentioned below to reach the consensus. Some of the classical consensus algorithms are proof of work and proof of stake, and some of the other algorithms were later on introduced to overcome challenges faced with the above-mentioned classical algorithms.

## *3.1 Proof of Work (PoW)*

Amongst many of the famous applications, Bitcoin is one of the applications which uses proof of work calculation to distribute a block as it plays out a great deal of computational work to demonstrate that the block is substantial or not. In proof of work, every node calculates the hash value of the given block with the continuously changing nonce value and the consensus decides whether the calculated value is less than or equal to the certain given value. If the calculated value satisfies the requirement criteria, then the given block is broadcast to all the other nodes and is considered valid and other nodes will append the validated block to their blockchains. The node which calculates the hash value is known as miners, and the process of calculating the hash value is known as mining.

In the decentralized network, as all the nodes are calculating the proof of work at the same time there is a possibility that more than one node generates a valid block at the same time, thereby creating a branch/fork as shown in Fig. 2, there is a chain of block-1, block-2, and block-3, and as you can see two valid blocks were generated simultaneously creating a branch or a fork. However, there is less probability that also the next block is calculated simultaneously; therefore, the longer chain will be considered now as a valid chain.

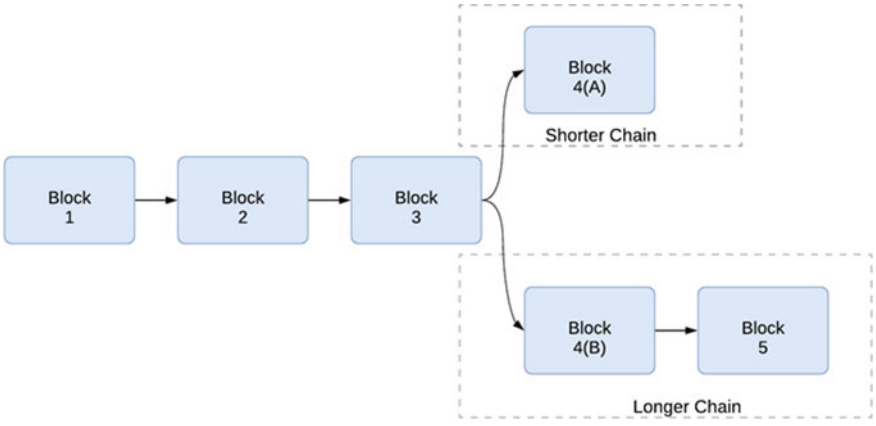



**Fig. 2** Fork effect while calculating proof of work

Miners need to go through a ton of PC counts to compute the proof of work; therefore, calculating proof of work results in a large waste of resources. Due to this high consumption of resources, a new cost-efficient consensus algorithm was introduced called proof of stake.

### *3.2 Proof of Stake (PoS)*

This consensus process can be considered equivalent to PoW, but instead of each node calculating the hash value of the changing nonce, here the nodes have to prove the amount of currency that they hold because it is believed that the person with a higher amount of currency is less likely to perform any malicious activity which is injustice. All things considered, the node with the most measure of cash would be ruling the organization. Subsequently, numerous different arrangements were proposed with the blend of proof of stake to choose which node will mine the following block. Some of the famous ones are:

- **BlackCoin**: BlackCoin uses a randomization method to decide which node mines the first block.
- **Peercoin**: Peercoin focuses on the age of the coin that the node with the oldest and a larger amount of coins is more likely to predict the next node.

### *3.3 Delegated Proof of Stake (DPoS)*

The working mechanism of DPoS is the same as PoS. To create the blocks, miners get priority based on their stake [1]. But the main difference between DPoS and PoS is that PoS is direct democratic while DPoS is representative democratic [1]. All the participant nodes select their representative node for the consensus process. With less amount of nodes for the validation process, blocks are confirmed quickly [5].

### *3.4 Ripple*

This algorithm uses a typical client–server architecture. Each node is partitioned into two types. Server for nodes takes interest in the agreement process and client for nodes participating in the exchange of assets. Each server has a unique node list (UNL) [4]. The main role of UNL is to decide where the node will be placed in the chain.

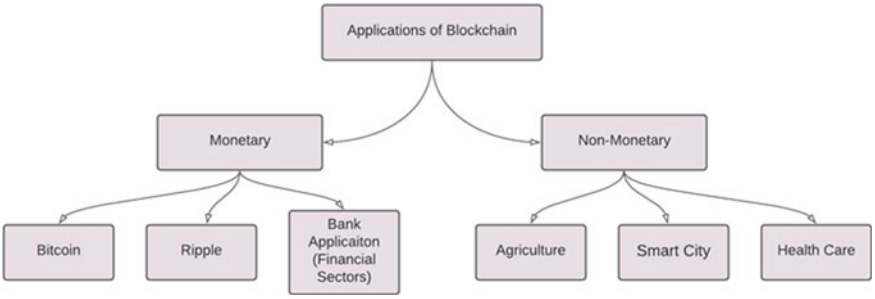



**Fig. 3** Classification of blockchain applications

### *3.5 Tendermint*

It is a type of Byzantine consensus algorithm. A random promoter is selected to broadcast the new unconfirmed block. The whole process is divided into 3 steps, i.e., (1) prevote step, (2) pre-commit step, and (3) commit step for validating newly arrived or created block

- **Prevote Step**: During this progression, validators conclude whether to create a prevote message for the new block.
- **Pre-commit Step**: If the proposed branch gets a lion's share vote in favor of the prevote, then it leads the branch to the pre-submit step. In the event that the block gets a lion's share of the decision in favor of pre-commit, therefore promote it to commit step.
- **Commit Step**: Nodes approve the block and broadcast it to all the blocks to accept the block if it receives a majority of votes [4]. This step provides the identity management of the connected object (Fig. 3).

## 4 Blockchain: Applications

### *4.1 Bitcoin*

Bitcoin is cryptographic cash planned in 2008 by a dark individual or group of people using the name Satoshi Nakamoto and started in 2009 when its execution was conveyed as open-source programming. It is decentralized automated money without a public bank or single chief that can be sent from customer to customer on the common Bitcoin network without the necessity for go-betweens [6]. Trades are checked by network center points through cryptography and recorded in a public circled record called a blockchain. Bitcoins are made as an honor for a cycle known as mining. They can be exchanged for various money-related norms, things, and organizations.

Bitcoin has been applauded and scrutinized. Pundits noticed its utilization in illicit exchanges, the enormous measure of power utilized by excavators, esteem unsteadiness, and robberies from exchanges. A couple of business examiners, including a couple of Nobel laureates, have depicted it as a hypothetical air pocket. Bitcoin has furthermore been used as an endeavor, but a couple of authoritative workplaces have given examiner cautions about Bitcoin.

### *4.2 Healthcare Sector*

Blockchain is currently in a very lead position in the biomedical field [7]. Further, it can be used to merge various fields of medicine such as genomics, telemedicine, neuroscience, and several others. Adaptability is given to the client to perform tasks with the medical services sensors, for example, utilizing a nasal wind stream sensor for perusing pace of wind current or ECG sensor for perusing pulse [8]. Blockchain can also be used in the hospital management system to manage the records of patients. Also, patients' records can be stored in a single database that is controlled or managed by a group of organizations. Blockchain can be also used as a firewall for managerial access to the main healthcare records. Bitcoin which is one of the applications of blockchain already displays the fundamentals of a trusted and auditable computing system used on a distributed system [9].

### *4.3 Ripple*

Ripple majorly focuses on credit networks [10]. It is divided into 6 parts which are listed as follows—server, ledger, last-closed ledger, open ledger, unique node list (UNL), and prosper [11]. It is an application which majorly focuses on banking markets. Ripple is open source; therefore, anybody with the relevant knowledge and skill set can use ripple for their applications [12].

### *4.4 Smart City*

An ideal concept of a smart city focuses on improving the quality of life of its citizens. This can be achieved with the integration of several leading technologies such as blockchain, Internet of things (IoT), and information and communication technology (ICT). Considering blockchain as a branch of ICT offering emerging opportunities ideally used for validating and maintaining cryptocurrency records but several studies later many immersive applications [13]. Thinking about the smart urban communities, it is available to dangers and the chance of dangers is delegated threats on accountability and threats on confidentiality [14].

### *4.5 Agriculture*

The worldwide food supply chain has been maintained by several participants such as farmers, retailers, transportation companies, wholesalers, and several others [15]. Maintenance of this chain is considered a very tedious task. This is where blockchain can be implemented and provide better functionality for the management of the whole food supply system. Agriculture and supply chain have been closely linked since the beginning as the agriculture industry also follows a similar structure. All the processes of the supply chain can be digitalized using various applications such as barcode, QR code, mobile application, and a digital certificate. Also, an advanced level of security can be provided using blockchain.

### *4.6 Financial Sector*

Capital business areas insinuate the coordinating of underwriters with enthusiasm for capital, to monetary masters with the contrasting threat, and bring profiles back. Whether or not the sponsor is business visionaries, new organizations, or colossal affiliations, the route toward raising capital can be trying. Firms face logically serious rules, longer events to get the occasion to grandstand, unusualness from advance expenses, and liquidity peril. Particularly in creating business areas, they ought to investigate the nonattendance of intensive checking, cautious rule, and sufficient market establishment for giving, settlement, clearing, and trading. Blockchain offers various favorable circumstances for a couple of capital market use cases which are recorded beneath:

- Elimination of a lone reason for dissatisfaction by displacing the rights from central utilities.
- Facilitation of capital market practices streamlining cycles, decreasing costs, and reducing time spent on settlements.
- Digitization of cycles and work measures, diminishing operational risks of blackmail, human misstep, and as a rule counterparty peril.
- Digitization or tokenization of favorable circumstances and budgetary instruments, making them programmable and significantly less complex to manage and trade. In the symbolic structure, they increment more broad market access through extended accessibility and the possibility of fractionalized ownership. These result in extended liquidity and reduced cost of capital (Table 3).

## 5 Blockchain: Challenges and Possible Solutions

There are several challenges faced during the implementation of blockchain technology. This section will discuss some of the general challenges faced during the

**Table 3** Recent trends of blockchain in several fields

| Paper | Published year | Field | Summary |
|---|---|---|---|
| [34] | 2020 | Energy sector | In this paper, the author played out a far-reaching audit of how blockchain innovation has been and can be conveyed in energy applications, going from the energy the executives to distributed exchanging to electric vehicle-related applications to carbon discharges exchanging and others |
| [35] | 2020 | Supply chain | In this paper, the authors have examined a great deal about the utilization of blockchain in the supply chain industry. Blockchain explicitly does not cover a great deal of this industry. Be that as it may, this paper examined a few practical choices in which blockchain can be utilized in SC applications |
| [36] | 2020 | Transport and logistics | This paper zeroes in an extraordinary arrangement on the utilization of blockchain in different applications of the transportation and logistics industry. It likewise examines conceivable future research accessible in this industry as in terms of blockchain |
| [37] | 2020 | Internet of things | This paper asserts that still there is a shortage of blockchain-based IoT applications accessible because of a few difficulties. In spite of the fact that the creators of this paper attempt to answer the greater part of these difficulties with practical arrangements |
| [38] | 2020 | Power systems | This article plans to propose a wide viewpoint about the use of blockchain innovation in the Power systems area, explaining some specialized perspectives concerning this promising innovation, the highlights, and applications grew up until now while zeroing in on the eventual fate of inventive applications in the electrical energy area |

integration of blockchain and later will discuss some of the issues while integrating blockchain with IoT and big data (Fig. 4).

**Anonymity** The main feature of blockchain is its transparency that each node can be audited at any point of time which causes an inverse effect on user privacy. Various blockchain applications lack this level of privacy [16]. Some of the attempts


**Fig. 4** Challenges faced by blockchain technology

to solve this issue were Zerocash [17] and Zerocoin [18] which provide a facility of a completely anonymous transaction system by encrypting the sender and receiver addresses [16]. Monero [19] uses a chain of signatures which makes transactions nearly impossible to trace back. As discussed earlier, there are various other applications such as Bitcoin fog [20], CoinJoin [21], Dark Wallet [22], and Dash [23] which was the first cryptocurrency that took anonymity and privacy into consideration.

**Security Threats** Fifty-one percentage of attack is the most common threat to blockchain [24]. Whenever any node gains control over the majority of the network, the node will now have the power to control the consensus process. Along with this, another major concern of malpractice is bribery [25]. Blockchain terminals or endpoints are where the clients communicate with the blockchain and it is likewise the weakest piece of the blockchain because of all the trading of information.

**Storage Capacity and Scalability** A large amount of storage is required in order to store the chain because a new block of sizes varying from various MBs is generated every few minutes. The increase in the size of the chain increases the number of resources needed to store the chain. There were various advancements taken into consideration, and some of them are Bitcoin-NG [26] and Litecoin [27] which is quite similar to Bitcoin but it provides improved confirmation time for transactions and also provides optimized storage.

**Smart Contracts** Basically, when the computer automatically does the work of a contract with the help of certain code it is known as smart contracts. It can be better understood as a smart contract code is implemented on the blockchain network which identifies each of the nodes/users of the transaction and helps the network in time efficiency cost reduction, precision, and transparency. This led to a newer version

of blockchain that is Blockchain 2.0. As it is a computer code, it is vulnerable to attacks such as hacking, viruses, and Trojans which harm the integrity of the software which in turn compromises the system. One of the examples of an application that has adopted this technology is Ethereum.

**Consensus** Any consensus algorithm is a backbone of the blockchain network as it performs a crucial task of whether to append the new block or not. The classic algorithm of consensus was proof of work (PoW) which was implemented in the first application of blockchain that is Bitcoin. Energy/resource consumption was one of the major limitations of PoW. Still, some of the cryptocurrencies such as Namecoin, Litecoin, Ethereum, and Bitcoin have still adopted this PoW as their consensus algorithms. There were several approaches to solve these limitations such as proof of stake (PoS), proof of burn (PoB) [28], proof of importance (PoI), proof of capacity (PoC), Paxos, and RAFT.

## *5.1 Challenges in IoT and Blockchain Integration*

Internet of things plays a vital role in building a modern environment around us by offering various smart solutions that are increasing our quality of living. IoT is the integration of software and hardware focused to make our everyday tasks easy. With all the recent advancements in this technology and providing a cloud platform for sharing a massive amount of data to and from it lacks some basics of data security. That is, the blockchain integrated with IoT can provide a better platform for the betterment of these technologies by providing the proper security, traceability, transparency, and reliability to the data. Integration of blockchain is IoT is not that hard but it also comes with its flaws which will be further discussed in this section:

**Storage Capacity and Scalability** As we have discussed earlier, the issue of storage is quite alarming but when integrated with IoT it becomes more complex as a typical IoT device can generate gigabytes and terabytes of data in a limited amount of time. Blockchain is not made for handling a large amount of data which could be an issue while implementing it with IoT applications.

**Security** Most of the IoT networks are prone to security attacks as they work on a wireless system. The main challenge with the integration of blockchain with IoT is the integrity of the data. If any corrupted data enters the blockchain, then it stays corrupted as the data once stored in the blockchain is immutable [16]. Some parts of the IoT network contain various physical devices that sometimes stop working properly or send malicious data which leads to security attacks on the network. Some of the programs such as GUITAR [29] and RemoWare [30] were introduced for run-time network updates which helped in keeping the smooth integration of blockchain and IoT.

**Anonymity and Data Privacy** Some of the IoT applications such as user health management, smart house, and various others generate very sensitive data that needs to be confidential. Securing this data, blockchain is quite tricky because one of the characteristics of blockchain is to be transparent. In order to make this integration more secure, several cryptographic algorithms should be implemented on the system to restrict unregistered access to the data in the system which is quite tricky.

**Smart Contracts** As discussed earlier, smart contract is one of the best applications of blockchain. But when it comes to the integration of blockchain and IoT, it gets a little trickier. A contract can be divided into two parts, one of which is code and another is data [16]. As IoT works on real-time data generation, accessing data from various sources would be complex for the contracts.

**Consensus** As IoT contains a greater number of physical devices that were designed to perform only the designated task, it would be very challenging to implement any consensus algorithm even as easy as proof of work or other algorithms. Keeping these challenges in mind, there were certain attempts to create an algorithm for consensus but still, the mining of the block would still be an issue. Proof of understanding (PoU) was an advanced version of PoW generated for this specific purpose of blockchain–IoT integration.

## *5.2 Challenges in Big Data and Blockchain Integration*

Big data is a field that focuses on the extraction of important data from a large data source such as a data warehouse or a large database. These data warehouses are filled with all the kinds of data collected from our smart devices, some of which are very sensitive such as user location and user information which if fallen into wrong hands is very disastrous. Blockchain can be integrated to this kind of application in order to monitor the flow of sensitive data and to protect the data from data breaches. In any such applications, blockchain acts as the mediator in order to restrict the access of service providers from gaining sensitive information from users. Let us take an example of a user-controlled medium that is Ushare [31] which gives users the rights to grant permission to the service provider for the access of its sensitive data as well as he/she can revoke those rights anytime they want. The ownership of data on the Internet is really flawed. Any data shared on the Internet is prone to be copied. The safety of these documents and artworks uploaded online is a very crucial task to do. But with blockchain technology, a digital registry can be generated for maintaining the content produced by the rightful owners and to find all the copies of any given document existing on the Internet and compare it to the original using various machine learning algorithms. One of the better examples of this technology is Ascribe [32] which was generated to carry out Internet document licensing. Ascribe uses its own protocol for blockchain implementation known as SPOOL which stands for Secure Public Online Ownership Ledger [33]. Several

artists face issues regarding their artworks being forged; for that reason, monograph was generated with the implementation of blockchain to provide a proper platform for buying and selling of artworks.

## 6 Conclusion

Blockchain holds the true potential to bring revolutionary changes to the field of technology. However, a portion of the impediments of blockchain is additionally talked about ensuing as (1) blockchain itself requires a great deal of time to measure and arrive at an agreement. In any case, when utilized with IoT, it is viewed as a weakness as time-taking applications are not proficient. (2) Blockchain utilizes an exorbitant measure of energy. (3) Scalability is a significant issue for blockchain applications. Likewise, examining future prospects or upgrades for blockchain as there are significant odds of the presence of cross-country acknowledged cryptographic forms of money or cryptocurrency. A few nations have proposed this thought. More integration with IoT will occur later on as a large portion of the leading organizations is looking for advances to this incorporated innovation. This paper has discussed the architecture of blockchain along with its types, several consensus algorithms, and some of the challenges faced by blockchain when integrated with several leading technological fields such as the Internet of things and big data for future development. Some of the existing solutions for the problems were also discussed.

## References

1. Zheng Z, Xie S, Dai HN, Chen X, Wang H (2018) Blockchain challenges and opportunities: a survey. Int J Web Grid Serv 14(4):352–375
2. Karafiloski E, Mishev A (2017) Blockchain solutions for big data challenges: a literature review. In: IEEE EUROCON 2017—17th ınternational conference on smart technologies. IEEE, pp 763–768
3. Basegio TL, Michelin RA, Zorzo AF, Bordini RH (2017) A decentralised approach to task allocation using blockchain. In: International workshop on engineering multi-agent systems. Springer, Cham, pp 75–91
4. Lin IC, Liao TC (2017) A survey of blockchain security issues and challenges. IJ Netw Sec 19(5):653–659
5. Bitshares—your share in the decentralized exchange. [Online]. Available https://bitshares.org/
6. Nakamoto S (2019) Bitcoin: a peer-to-peer electronic cash system. Manubot
7. Kuo TT, Kim HE, Ohno-Machado L (2017) Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 24(6):1211–1220
8. Wang Haoxiang (2020) IoT based clinical sensor data management and transfer using blockchain technology. J ISMAC 2(03):154–159
9. Linn LA, Koo MB (2016) Blockchain for health data and its potential use in health it and health care related research. In: ONC/NIST use of blockchain for healthcare and research workshop. ONC/NIST, Gaithersburg, Maryland, United States, pp 1–10

10. Ghosh A, Mahdian M, Reeves DM, Pennock DM, Fugger R (2007) Mechanism design on trust networks. In: International workshop on web and ınternet economics. Springer, Berlin, Heidelberg, pp 257–268
11. Schwartz D, Youngs N, Britto A (2014) The ripple protocol consensus algorithm, vol 5(8). Ripple Labs Inc White Paper
12. Armknecht F, Karame GO, Mandal A, Youssef F, Zenner E (2015) Ripple: overview and outlook. In: International conference on trust and trustworthy computing. Springer, Cham, pp 163–180
13. Pieroni A, Scarpato N, Di Nunzio L, Fallucchi F, Raso M (2018) Smarter city: smart energy grid based on blockchain technology. Int J Adv Sci Eng Inf Technol 8(1):298–306
14. Vivekanadam B (2020) Analysis of recent trend and applications in block chain technology. J ISMAC 2(04):200–206
15. Kamilaris A, Fonts A, Prenafeta-Boldú FX (2019) The rise of blockchain technology in agriculture and food supply chains. Trends Food Sci Technol 91:640–652
16. Reyna A, Martín C, Chen J, Soler E, Díaz M (2018) On blockchain and its integration with IoT. Challenges and opportunities. Future Gener Comput Syst 88:173–190
17. Sasson EB, Chiesa A, Garman C, Green M, Miers I, Tromer E, Virza M (2014) Zerocash: decentralized anonymous payments from bitcoin. In: 2014 IEEE symposium on security and privacy. IEEE, pp 459–474
18. Miers I, Garman C, Green M, Rubin AD (2013) Zerocoin: anonymous distributed e-cash from bitcoin. In: 2013 IEEE symposium on security and privacy. IEEE, pp 397–411
19. Monero 2017. https://getmonero.org/. Accessed 20 October 2017
20. Bitcoin Fog, 2017. Available online http://bitcoinfog.info/. Accessed 1 February 2018
21. Maxwell G (2013) CoinJoin: Bitcoin privacy for the real world. In: Post on Bitcoin forum
22. Greenberg A (2014) 'Dark Wallet'is about to make Bitcoin money laundering easier than ever. URL http://www.wired.com/2014/04/dark-wallet
23. Dash, 2017. https://www.dash.org/es/. Accessed 20 October 2017
24. Eyal I, Sirer EG (2014) Majority is not enough: Bitcoin mining is vulnerable. In: International conference on financial cryptography and data security. Springer, Berlin, Heidelberg, pp 436–454
25. Bonneau J, Felten EW, Goldfeder S, Kroll JA, Narayanan A (2016) Why buy when you can rent? Bribery attacks on Bitcoin consensus
26. Eyal I, Gencer AE, Sirer EG, Van Renesse R (2016) Bitcoinng: a scalable block-chain protocol. In 13th (USENIX) symposium on networked systems design and implementation (NSDI 16), pp 45–59
27. Litecoin, 2011. https://litecoin.org/. Accessed 4 February 2018
28. Stewart I (2012) Proof of burn. bitcoin. it. Available online https://en.bitcoin.it/wiki/Proof_of_burn. Accessed 4 March 2018
29. Ruckebusch P, De Poorter E, Fortuna C, Moerman I (2016) Gitar: generic extension for Internet-of-Things architectures enabling dynamic updates of network and application modules. Ad Hoc Netw 36:127–151
30. Taherkordi A, Loiret F, Rouvoy R, Eliassen F (2013) Optimizing sensor network reprogramming via in situ reconfigurable components. ACM Trans Sens Netw (TOSN) 9(2):1–33
31. Chakravorty A, Rong C (2017) Ushare: user controlled social media based on blockchain. In: Proceedings of the 11th international conference on ubiquitous information management and communication, pp 1–6
32. McConaghy T, Holtzman D (2015) Towards an ownership layer for the internet. ascribe GmbH
33. de Jonghe D (2016) SPOOL Protocol. https://github.com/ascribe/spool
34. Bao J, He D, Luo M, Choo KKR (2020) A survey of blockchain applications in the energy sector. IEEE Syst J
35. Durach CF, Blesik T, von Düring M, Bick M (2020) Blockchain applications in supply chain transactions. J Bus Logistics
36. Pournader M, Shi Y, Seuring S, Koh SL (2020) Blockchain applications in supply chains, transport and logistics: a systematic review of the literature. Int J Prod Res 58(7):2063–2081

37. Rao AR, Clarke D (2020) Perspectives on emerging directions in using IoT devices in blockchain applications. Internet Things 10:
38. Di Silvestre ML, Gallo P, Guerrero JM, Musca R, Sanseverino ER, Sciumè G, Vásquez JC, Zizzo G (2020) Blockchain for power systems: current trends and future applications. Renew Sustain Energ Rev 119:

# Filter Bank Multicarrier Systems Using Gaussian Pulse-Based Filter Design for 5G Technologies


**Deepak Singh and Mukesh Yadav**

**Abstract** FBMC has arisen the most multicarrier (MC) procedure since it satisfies the need of the next-generation 5G standards. Optimized generalized Gaussian pulse (OGGP) filter of the FBMC systems have been proposed in this paper. A PTF strategy based on optimized generalized Gaussian pulse is proposed along with the lowest time domain side lobe levels and tail energy. The proposed system has been used to OQAM modulation and different Tx and Rx antenna. From the numerical perspective, the transmission execution can be improved by considering the greatest signal-to-noise ratio and reduced ISI. The proposed system is simulated by using MATLAB software and also by using the BER and magnitude response.

**Keyword** Filter bank multicarrier (FBMC) · OGGP · MIMO system · 5G technology

## 1 Introduction

The flow research interest on mobile communication system is going through a change in a perspective from the generally settled third and fourth era advancements, to a much adaptable future 5G networks [1]. 5G networks, which are required to be sent in 2020, target giving its clients with giga-bit information rate insight under expanding information traffic-request and high portability, high throughput alongside low inactivity, productive use of accessible range, and so forth, at decreased expense and with low force utilization. In any case, customary OFDM experiences different disadvantages, which settle on it a restricted decision to meet the assorted and testing necessities of the future cell frameworks. To battle time scattering and to encourage per sub-transporter balance, OFDM utilizes an excess time stretch called cyclic prefix, between two images [2].

The presence of cyclic prefix prompts wasteful utilization of the accessible transmission time, there by upsetting the low idleness prerequisite, prompting low ghostly

D. Singh (✉) · M. Yadav
Department of Electronics and Communication, SIRT, Bhopal, MP, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_45

productivity, and at last, diminished throughput. The communicated and get channel utilized in OFDM is time-restricted rectangular heartbeat, which has high side-projections in the recurrence area that makes phantom spillage neighboring sub-transporters. These out-of-band emanations eventually lead to ICI. The presence of huge side-projections additionally restricts its application to intellectual radio situation. The symmetry among the sub-transporters is misshaped under quick blurring versatile channels which likewise incites ICI. Under doubly dispersive channels including both time and recurrence spreading, OFDM experiences both ISI and ICI, which requests rigid synchronization procedures, which builds the force utilization.

In FBMC, the out-of-band emanations are diminished by utilizing adaptable heartbeat molding channels which are all around confined in time and recurrence spaces, thereby giving vigor against time-recurrence scatterings brought about by doubly dispersive channels [3, 4]. FBMC frameworks, particularly OFDM-OQAM, can accomplish upgraded ghostly productivity and decreased ISI/ICI without the presence of cyclic prefix. In FBMC, a bunch of information images, which should be communicated at a specific sub-transporter, is gone through a communicate channel adjusted at that relating sub-transporter. The information adjusted at various sub-transporters are totaled and also sent through the channel.

At the collector, the information images are isolated out by an examination channel bank. In OFDM-OQAM, not at all like the cyclic prefix-OFDM, the complex information images are sent as genuine images, for example, the in-stage and quadrature-stage segments are communicated at each time moments that are products of half of the image span, thereby keeping up great ghastly effectiveness, even with no cyclic prefix. The usage parts of the various variations of FBMC frameworks through productive equipment models and polyphase structures are concentrated [5]. It is significant that the model heartbeat forming channel utilized at the examination and amalgamation channel banks assumes a critical function in defeating the restrictions of OFDM.

## 2 FBMC Technique

The Tx/Rx block outline for an overall symmetrical multicarrier balance framework is introduced in Fig. 1. Despite the fact that Fig. 1 can be relevant to both FBMC and OFDM, the rectangular channel is the model channel utilized in OFDM, whereas different strong channels serve the capacity of model channel in the last case. In the event that the information image to be sent at the $n$-th time moment and $k$-th subcarrier is indicated as $d_k(n)$, the communicate signal is created as

$$s(t) = \sum_{n} \sum_{k=0}^{N-1} d_k(n) h(t - nT) \mathrm{e}^{j2\pi(t-nT)kF} \tag{1}$$

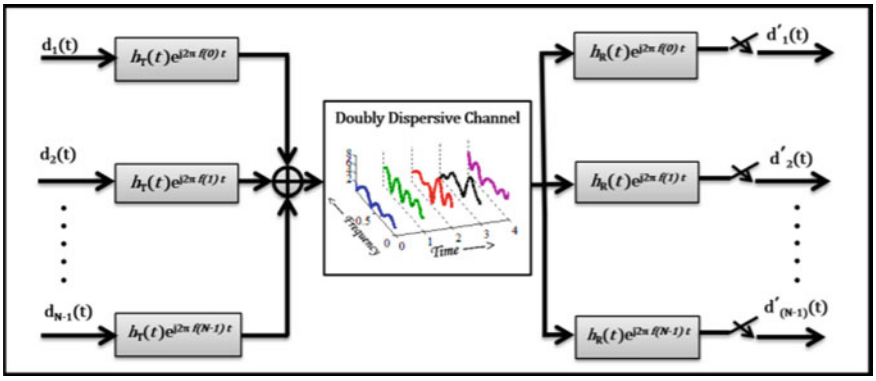



**Fig. 1** Block diagram for a multicarrier modulation system

Here, $T$ and $F$ mean the time separating between progressive images and the sub-transporter separating individually. $N$ means the quantity of sub-transporters, and $h(t)$ speaks to the motivation reaction of the communicate model channel utilized in FBMC framework. The information images $d_k(n)$ are spread over the time-recurrence lattice as clarified [6–10].

In OFDM-OQAM frameworks, the in-stage and quadrature-stage parts of the perplexing images $d_k(n)$ are stumbled in time space by a factor $T/2$, so that it can be considered as two heartbeat adequacy balanced symbols sent with a timing balance of $T/2$ [11–15].

Let $d_k(n)$ be written as

$$d_k(n) = d_k^I(n) + jd_k^Q(n) \tag{2}$$

At the collector, the got signal is passed through a coordinated channel bank and is examined at the image time dispersing, $T$, so that the sent images can be demodulated. Under ideal channel conditions, the sent information images are recoverable at the recipient if the model channel fulfills certain conditions which are summed up. Let $k$ and $I$ indicate the sub-transporter lists at the transmitter and collector individually and the comparing image testing time files be $m$ and $n$.

## 3 OGGP Based Filter

The conventional filter issues were focused on planning filter that are isotropic and with better recurrence reaction attributes to decrease the out-of-band radiation. The proposed OGGP based filter is dissected and contrasted and the regular filter utilizing shape plots of equivocalness capacity, time and recurrence space side-projections and tail energy, and so on. The proposed filters have lower side-flaps and quicker rot in

the time area contrasted with the other existing filters. A broader plan technique for the plan of OGGP channel is likewise proposed which prepares to improve the time-space qualities of the channel by fluctuating only one plan boundary. The equation of OGGP based filter is given by:

$$g_{\alpha,T}(t) = \frac{1}{1-\alpha}\left(e^{-4\pi\left(\frac{t}{t}\right)^2}\right) - \alpha\left(e^{-4\pi\left(\frac{\alpha t}{t}\right)^2}\right) \tag{3}$$

$\alpha$ is the scaling boundary which chooses the recurrence confinement, and $T$ is the ordinary term. The GGP will have a zero intersection in time area as indicated by estimations of $\alpha$ and $T$. Thus, a channel symmetrical in time area that fulfills Nyquist standard can be developed if the plan boundaries $\alpha$ and $T$ are fittingly picked, so that the vagueness capacity will have zero intersection at the ideal focuses.

## 4 Proposed Methodology

The articulation for summed up OGGP based filter is yielded (4). Here, $\alpha$ is a scaling boundary, and $T$ speaks to the ostensible span of the OGGP. In OGGP plan issue, the boundary $T$ is gotten by fixing α and discovering the zero intersection of the beat by likening the (4) to zero. In OGGP plan, a connection between the zero intersection of the beat and the equivocalness work as expected space is assessed by bend fitting, and the estimation of $T$ is gotten for the given $\alpha$ esteem. This methodology is heuristic since the proportionality consistent relies upon the $\alpha$ worth and it changes for various $\alpha$ values. Henceforth, an overall methodology to discover the ideal $T$ should be planned. The uncertainty work along time hub is same as the autocorrelation of the beat. Consequently as opposed to discovering $T$ from the equivocalness work, the autocorrelation of the GGP channel can be utilized. The autocorrelation of $h_{\alpha,T}(t)$ is gotten as

$$x(t) = \int_{-\infty}^{\infty} h_{\alpha,T}(\rho) h_{\alpha,T}(t+\rho) d\rho \tag{4}$$

It could be seen that the SIR amplifying condition relies upon the proportion of defer spread and Doppler spread of the doubly dispersive channel instead of the specific estimations of the spread which is shown in Fig. 2.

It could be seen that the SIR amplifying condition relies upon the proportion of defer spread and Doppler spread of the doubly dispersive channel instead of the specific estimations of the spread as shown in Fig. 2. The underlying loads are admirably picked by the standardized postponement and Doppler spreads with the goal that the equivocalness work spreads across time and recurrence tomahawks with the end goal that the SIR is augmented. The extended inclination calculation

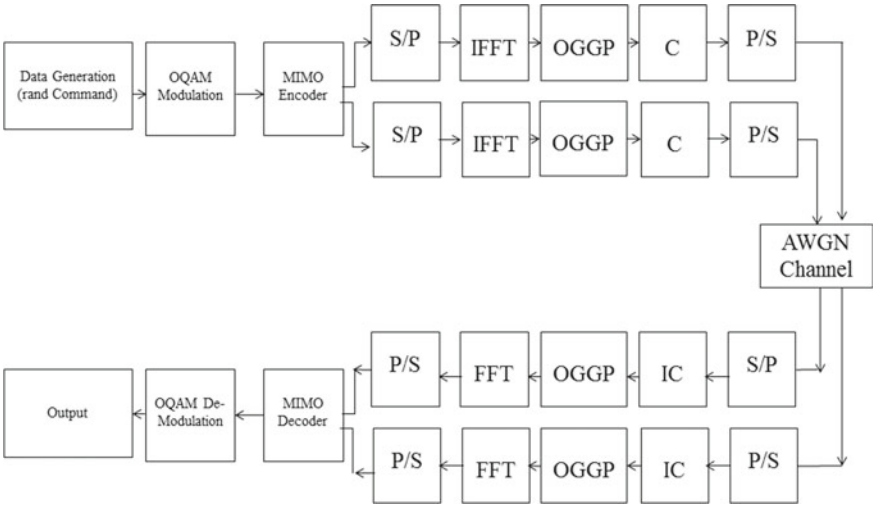



**Fig. 2** FBMC transceiver using OGGP

is applied with these loads as starting worth, so that the vagueness work spread is kept up. In the interim, the extent of vagueness work at those focuses, where a base worth is required, is pulled down to zero. Consequently, for a specific deferral and Doppler spread, the vagueness capacity of the planned heartbeat has a shape which amplifies SIR and simultaneously fulfills the Nyquist rules as expected and recurrence domains. The key thought of the part is to adaptively shift the plentifulness scaling boundary, so the beat accomplishes a shape which amplifies the SIR. The numerical articulation for the time-space and recurrence area side-flap energy is likewise inferred in the current investigation. In light of the greatest reasonable side-projection energy as expected and recurrence spaces, the upper and lower cutoff points of the adequacy scaling boundary can be fixed. The connection between the time-recurrence scattering of the beat and the adequacy scaling boundary is too set up. A shut structure articulation for the last heartbeat shape is additionally inferred in the present commitment utilizing Jacobian theta work. The adjustment fits as a fiddle of the channel, and furthermore, its vagueness work for various estimations of the scaling boundary and the precision of the detailing of the shut structure articulation for the model heartbeat are broke down through reenactments. The SIR execution of the proposed strategy under doubly dispersive channels is dissected and contrasted, and the traditional isotropic model channel plans.

## 5 Simulation Results

The OGGP channel is intended for various estimations of $\alpha$ as clarified is area. The standardized image timing span and subcarrier transmission capacity are taken as $\sqrt{2}$. The various estimations of $\alpha$ and the comparing $T$ esteems got utilizing the proposed strategy are organized. The assembly plots of $T$ versus number of emphases are shown in Figs. 3, 4, and 5.

The size of mistake versus emphasis number for different $\alpha$ values is appeared in Fig. 4. It very well might be noticed that the plots meet inside eight emphases. There is a slight expansion in the quantity of cycles for expanding $\alpha$ be that as it may, which is unimportant.

The various channels and their uncertainty capacities utilized for creating the ideal heartbeat are delineated. Figure 5 speaks to the rectangular, square root raised cosine, and Gaussian heartbeats individually. The form plots of their uncertainty capacities are shown in Fig. 6 individually. The shape plot of the DAF of rectangular channel is delineated.

In the shape plots, the sizes at different network focuses are shown by various tones. The planning between the greatness esteems and the shading can be acquired from the shading bar close to the form in Fig. 7.

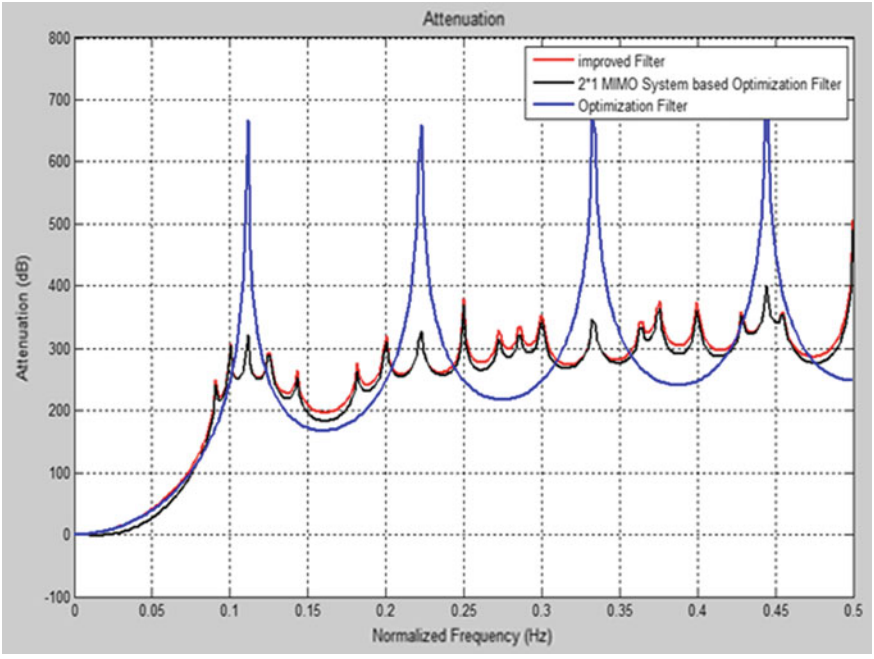



**Fig. 3** Attenuation of 2 × 1 system-based OGGP based filter

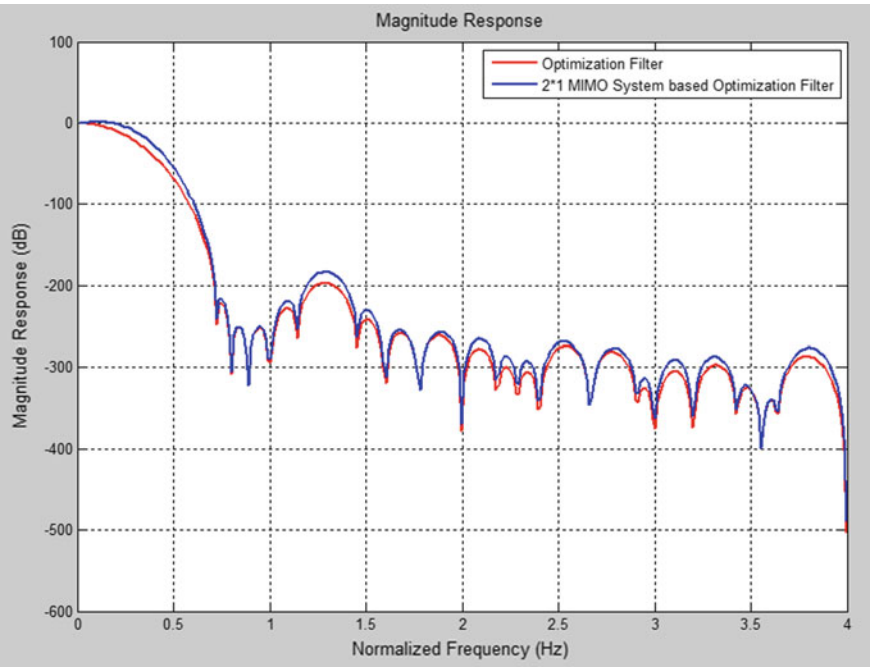



**Fig. 4** Magnitude response of 2 × 1 system-based OGGP based filter

White regions speak to an extent more modest than the base worth showed in the shading bar. The DAF of rectangular channel is spread along the vertical hub since the beat is not band restricted. Since the rectangular heartbeat becomes beat in the phantom area shown in Fig. 8, the side-flap sizes fall in the range of −20 to −13 dB along the recurrence hub.

## 6 Conclusion

PTF plan strategies for FBMC frameworks and visually impaired assessment techniques with OFDM were proposed in this postulation. Two filters, which have fixed time/frequency area attributes and two versatile channels, whose time/frequency area attributes differs as per channel varieties, are planned. The planned channels are fundamentally obtained from the ideally restricted Gaussian channel and have subsequently shown great time/frequency space restriction. The plan method of the channels has guaranteed that they will fulfill Nyquist models in both time and frequency spaces, so that the images sent are entirely reproduced without the channel.

The OGGP filter displays the most reduced time-space side-flap levels among the different traditional and proposed fixed channels.

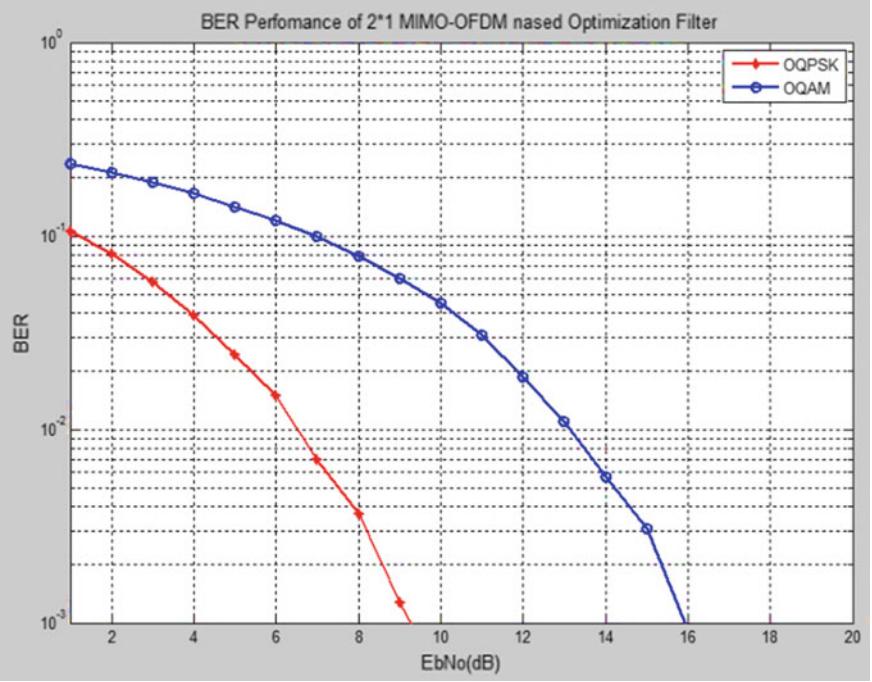



**Fig. 5** BER of 2 × 1 system-based OGGP based filter

The proposed work has the state of a sinusoid and henceforth assessed utilizing just three preliminary qualities. The presentation improvement of the proposed strategy is appeared under doubly dispersive channels with huge defer spreads and also Doppler spreads.

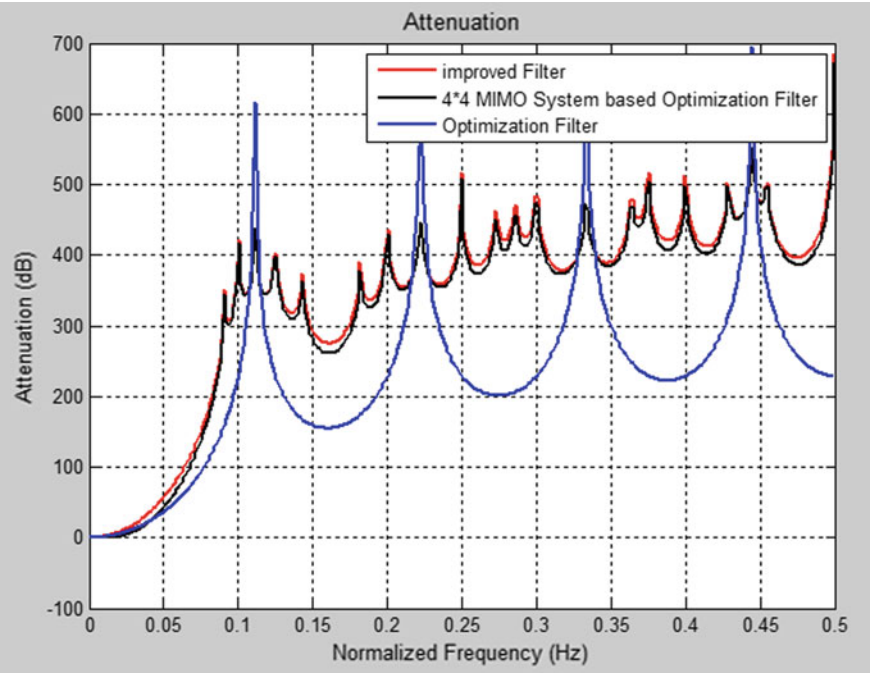



**Fig. 6** Attenuation of 4 × 4 system-based OGGP based filter

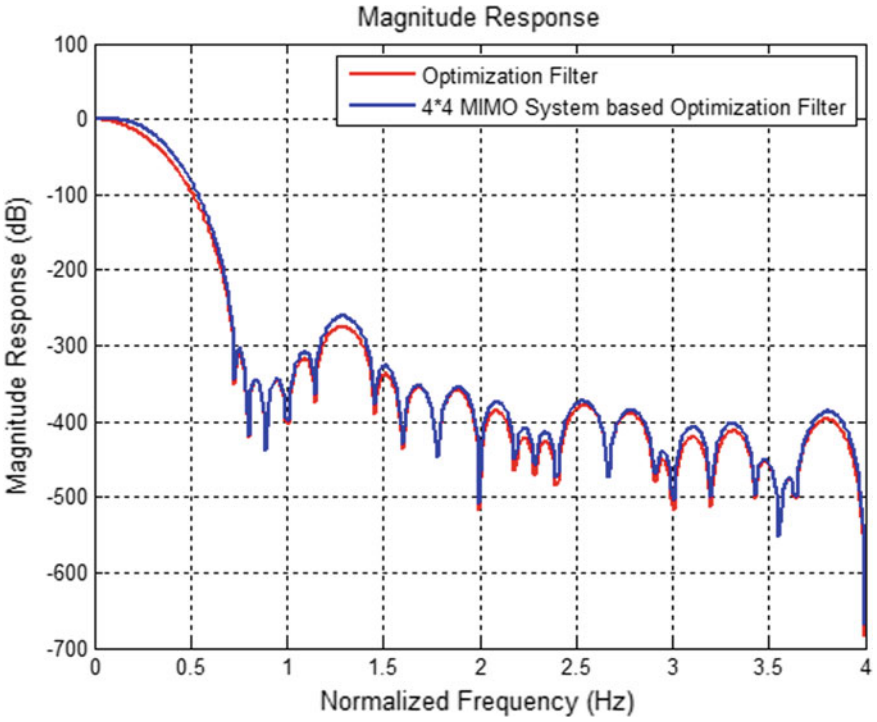



**Fig. 7** Magnitude response of 4 × 4 system-based OGGP based filter

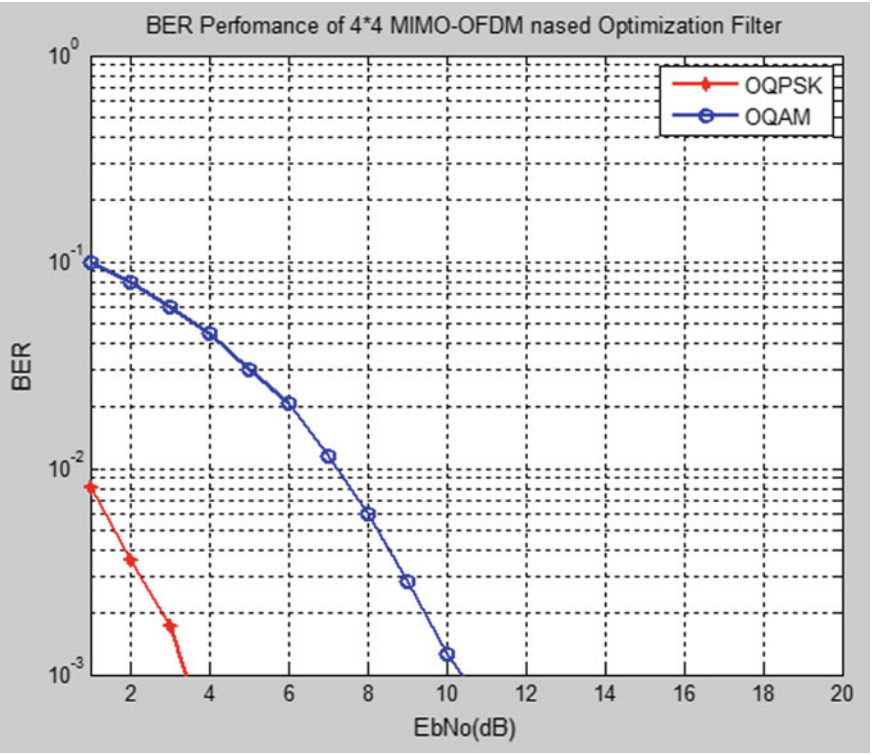



**Fig. 8** BER of 2 × 1 system-based OGGP based filter

## References

1. Jeffery JMJ, Masapalli L, Nookala VM, Dasari SP, Kirthiga S (2020) Peak to average power ratio and bit error rate analysis of multi carrier modulation techniques. In: International conference on communication and signal processing, pp 1443–1446
2. Rachini SA, Jaber MM (2019) Performance of FBMC in 5G mobile communications over different modulation techniques. In: International symposium on networks, computers and communications (ISNCC), 01–06, 2019
3. Nissel R, Ademaj F, Rupp M (2018) Doubly-selective channel estimation in FBMC-OQAM and OFDM systems. In: IEEE 88th vehicular technology conference (VTC-Fall), 01–05, 2018
4. He Z, Zhou L, Chen Y, Ling X (2017) Filter optimization of out-of-band emission and BER analysis for FBMC-OQAM system in 5G. In: 9th IEEE ınternational conference on communication software and networks, pp 34–39
5. Mawlawi B, Dore JB, Berg V (2015) Optimizing contention based access methods for FBMC waveforms. In: International conference on military communication and ınformation systems, 01–06, 2015
6. Park H, Park SH, Kong HB, Lee I (2012) Weighted sum MSE minimization under PerBS power constraint for network MIMO systems. IEEE Commun Lett 16(3):360–363
7. Farhang-Boroujeny B (2011) OFDM versus filter bank multicarrier. IEEE Sig Process Mag 28:92–112

8. Chen D, Qu DM, Jiang T (2010) Novel prototype filter design for FBMC base cognitive radio systems through direct optimization of filter coefficients. In: IEEE ınternatioanl conference wireless communication and signal processing, pp 21–23
9. Farhang-Boroujeny B (2010) Cosine modulated and offset QAM filter bank multicarrier techniques a continuous-time prospect. EURASIP J Adv Sig Process 1–6
10. Viholainen A, Ihalainen T, Stitz TH, Renfors M, Bellanger (2009) Prototype filter design for filter bank based multicarrier transmission. In: 17th European signal process conference, pp 24–28
11. Ari V, Maurice B, Mathieu H (2009) WP5: prototype filter and filter bank structure. In: Phydyasphysical layer for dynamic access and cognitive radio
12. Ikhlef A, Louveaux J (2009) An enhanced MMSE per subchannel equalizer for highly frequency selective channels for FBMC/OQAM systems. In: Processing IEEE 10th workshop on signal processing advances in wireless communications, pp 186–190
13. Waldhauser DS, Baltar LG, Nossek JA (2008) MMSE subcarrier for filter bank based multi-carrier systems. In: Processing IEEE 9th workshop on signal processing advances in wireless communications, pp 525–529
14. Kim I, Park IS, Lee YH (2006) Use of linear programming for dynamic subcarrier and bit allocation in multiuser OFDM. IEEE Trans Veh Technol 55(4):1195–1207
15. Siohan P, Siclet C, Lacaille N (2002) Analysis and design of OFDM/OQAM systems based on filter bank theory. IEEE Trans Sig Process 50(5):1170–1183

# LIMES: Logic Locking on Interleaved Memory for Enhanced Security


**A. Sai Prasanna, J. Tejeswini, and N. Mohankumar**

**Abstract** Globalization and increasing distribution of IC supply chain have resulted in various third parties having a key to precious intellectual property or the physical integrated circuit and therefore information can be exploited. Information security is the practice of safeguarding information by minimizing information risks. It is required to scale down the danger of unauthorized information disclosure, modification, and destruction. Hardware security threats have been observed at several levels of the IC supply chain. To protect the hardware from potential attacks, there are various design-for-security (DFS) techniques. Interleaved memory with logic locking techniques for information security is proposed in this paper. Interleaved memory is a solution for random arrangement and logic locking is needed to interrupt the chain and to protect the data by restricting its access to authorized users. It is a versatile and easy-to-integrate solution that needs only trusted designers. The approach proposed in this paper compares the Hamming distance (HD) and Levenshtein distance (LD) and BER obtained for random logic locking (RLL) and weighted logic locking (WLL) on the interleaved memory. From the results obtained, it can be concluded that weighted logic locking on interleaved memory provides better security.

**Keywords** Hardware security · Interleaved memory · Logic locking · Design for security · Information security

A. Sai Prasanna · J. Tejeswini · N. Mohankumar (✉)
Department of Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: n_mohankumar@cb.amrita.edu

A. Sai Prasanna
e-mail: sairam1199@gmail.com

J. Tejeswini
e-mail: tejes748@gmail.com



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_46

## 1 Introduction

Information security is the use of physical and technical controls in an endeavor to shield the confidentiality, integrity, and availability of data. It is needed to cut back the danger of unauthorized information leaks, modification, and deletion. Since the advent of the Internet, cybersecurity has been a significant research and business focus. Since an attack can occur on any of the layers of the cyberenvironment (e.g., operating system, network, and application software), there are several strategies to tackle different cybersecurity threats, including malware, ransomware, denial of service, etc. Usually, a hardware attack results in more serious and hard-to-recover damage. Also, low-cost attacks based on hardware may be devastating. Hence, this paper investigates issues related to hardware security. Hardware used for various applications from customer electronics to defense systems requires a secured access, as there emerges various threats in the integrated circuit (IC) design flow [1]. A hardware-based attack occurs as a result of an attacker having the ability to take advantage of hardware vulnerabilities to cause a trade-off in the system security [2]. Outsourcing of the fabless IC manufacturing companies to the untrusted foundries leads the major cause of these threats [1]. The different threats are reverse engineering [3, 4], IP piracy [4, 5], overbuilding [6, 7], hardware Trojans [8, 9], and counterfeiting [10, 11]. Solely due to the intellectual property right violations, the semiconductor industry losses about $4 billion per annum [6, 11]. Information or data is stored in the memory based on the memory address assigned to it. The address assigned can be either in order or randomly arranged. Generally, the orderly arrangement of address takes place. But orderly arrangement of data makes it easily accessible whereas random arrangement like in interleaved memory is a little tricky. To protect this data, design-for-security techniques are to be used such as watermarking [8, 10], fingerprinting [11], IC metering [12], split manufacturing [5, 13], and logic locking (LL) [1, 14], but logic locking is one technique that has proved to be resilient to maximum attacks and is easy to incorporate in applications.

Logic locking hides features of the structure and can be opened with a secret key. Owing to its simplicity, logic locking has gained notable recognition from the research community [6].

The remainder of this paper is organized as follows. Section 2 talks about interleaved memory for information security. Section 3 discusses the threats faced to keep information secure, sheds light on various designs-for-trust (security) approaches and how logical locking is more efficient than the other methods. Section 4 deals with logic locking and different ways to implement logic locking and threats faced by it. Section 5 describes the proposed method. Section 6 performs analysis of the methods discussed. Concluding remark is given in Sect. 7.

## 2 Interleaved Memory

Interleaving is the method of reorganizing the positioning of a data sequence to form a rearranged version of its initial self and is carried out at the transmitter side and the converse of this operation is deinterleaving which restores the obtained sequence to the initial order observed at the receiver side [2].

Error correction coding is believed to be a fruitful way to cope with various types of error in data communication. Interleaving is conventionally suggested to minimize bit error rate (BER) of digital transmission over a channel with data transmitted as bursts of information. The interleaving architecture enhances the performance at the expense of complexity, memory requirement, and time delay [13].

There are two conventional methods of interleaving, namely the block interleaver and convolutional interleaver. In the block interleaver, the input data is noted rowwise into a memory that is established as a matrix and retrieved columnwise. In the convolutional interleaver structure, the information is multiplexed through and from a fixed amount of shift registers [2].

Generally, the orderly arrangement of address takes place. But orderly arrangement of data makes it easily accessible whereas random arrangement like in interleaved memory is a little tricky. Therefore, the interleaved memory approach to storing data is better at securing information, yet this memory is not completely fool proof and is vulnerable to attacks.

## 3 Threats and Design for Security

To avoid any inside or outside danger such as criminals or any person who aims to impede or undermine the organization's sustainable state, security measures are taken. The protection measures taken against the damages that may happen are known as the safety [15]. Hardware used for various applications from customer electronics to defense system requires a secured locking, as there emerges various threats in the integrated circuit (IC) design flow [1]. The various threats are (i) reverse engineering [3, 12]. (ii) IP piracy [12]. (iii) Overbuilding [12]. (iv) Hardware Trojans [9]. (v) Counterfeit ICs [10, 12].

### *3.1 Design for Security*

Passive (watermarking, finger printing, etc.) and active (logic locking) techniques are the different types of classification of hardware security techniques [16].

- Watermarking—unnoticeably modifying a part of information. Helps in the detection of piracy/theft after its occurrence but no prevention.

- Fingerprinting—signatures of designer and end users are embedded together. Similar to watermarking only detection.
- IC metering—unique ID assigned to ICs being manufactured to track them.
- IC camouflaging—masking chosen gates. Protection against unreliable end users only [12].
- Split manufacturing—splitting the process of manufacturing and then combining together. Protection against unreliable foundries alone [5].
- P. kumar et al. proposed a toggle count-based node selection for module placement to enhance security [17].
- Logic locking—introducing an additional hardware segment into a design to lock/secure the design [12].

### *3.2 Choice of Logic Locking*

The most sought out design-for-trust (DfTr) technique is logic locking. Without knowing the correct key, even if an adversary manages to lay their hands on the encrypted netlist, it will not be of any use as the correct functional design cannot be obtained [6]. With logic locking in place, overbuilding by the foundry is minimized as the excess ICs cannot be activated [6]. Trojan insertion is prevented as the attacker finds it difficult to find suitable location to place them [12].

## 4 Logic Locking

In this technique, additional logic is inserted into a circuit and the primary design is locked with a secret or confidential key. The correct output is produced only when correct key values are applied [7, 18]. Logic locking can be implemented in two ways: *sequential logic locking* [5, 8, 11] and *combinational logic locking* [4, 7, 11]. The two logic locking techniques and different types of combinational logic locking are as shown in Fig. 1.

- Random logic locking (RLL) or EPIC—instituted by Roy et al. [14] in 2008. XOR key gates are placed at arbitrary locations in the netlist [6]. Suffers from low HD. Prone to key-sensitization attack.
- Fault analysis LL—nodes with high fault impact are identified and key gates are introduced at these nodes [1]. Provides 50% HD but results in longer implementation time. Prone to key-sensitization attack [16].
- Strong LL—Key gates with maximum mutual interference are incorporated [11]. Resistant to key-sensitization attack. But it provides low HD rates [16] and insertion of pairwise secure key gates depends highly on the circuit topology [11].
- Weighted LL—It controls the design with multiple key inputs and offers:

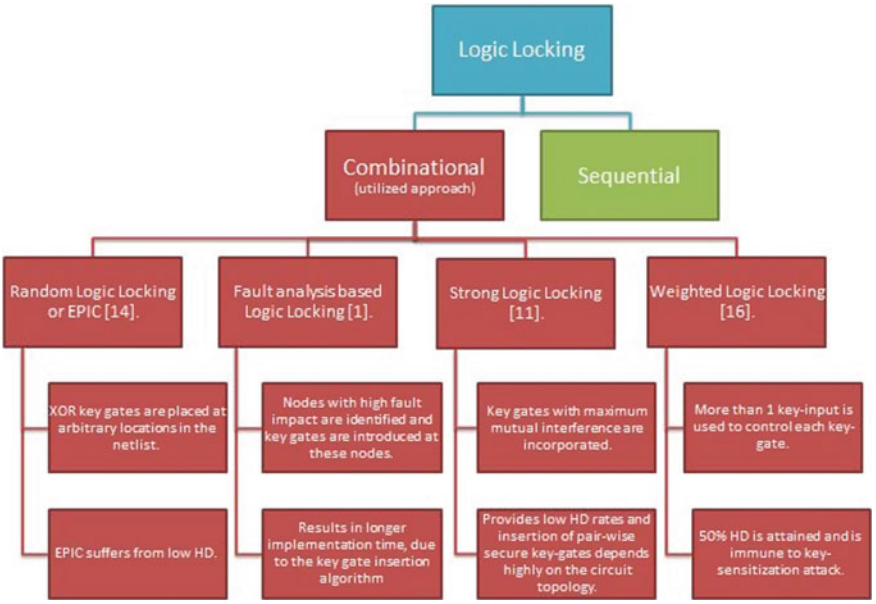



**Fig. 1** Block diagram depicting different logic locking techniques

- Immune to key—sensitization attack
- 50% HD due to an efficient key gate insertion metric, and
- Notably refined output corruption efficiency of the key gates resulting in a much shorter execution time [16].

## 5 Proposed Method

This paper proposes the idea of protection of data that is first stored in an interleaved memory and further uses logic locking technique to enhance security. Block diagram representing the proposed method is shown in Fig. 2.

In an interleaved memory organization, the data is either evenly arranged in consecutive address blocks or can be arranged randomly using linked lists or hash table data structure. This random arrangement of data provides a layer of security than the serially arranged data. This interleaved data is protected using logic locking techniques. Here, it has been proposed to use both RLL and WLL and compare their efficiencies. As shown in Fig. 3, the input data is sent to an interleaver which scrambles the data and either RLL or WLL is implemented on this data. The HD%, LD and BER calculation are performed and these responses' values are checked. If the values are not close to the optimal result, then the design is modified to add more logic gates to increase security and the output is taken as the final output.

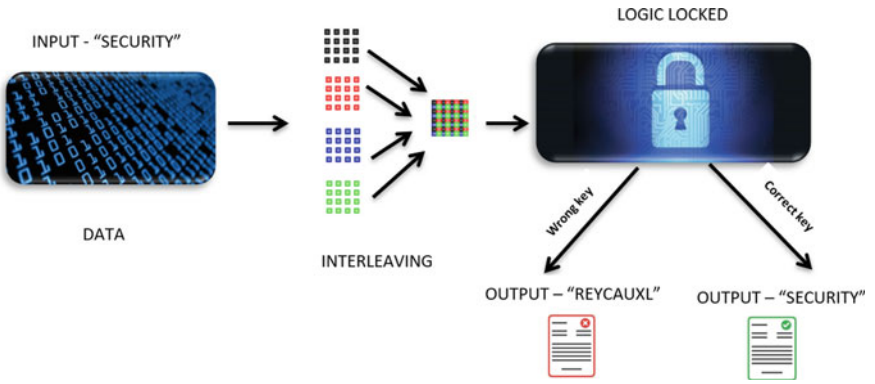



**Fig. 2** Block diagram representation of proposed method

Without knowing the correct key, even if an adversary manages to lay their hands on the encrypted netlist, it will not be of any use as the correct functional design cannot be obtained [6]. It hides some part of the functionality in the secret key bits; the structural information obtained from hardware implementation reveals only the remaining part of the functionality [9]. LL protects the system from IC piracy and reverse engineering. Overbuilding by the foundry is minimized as the excess ICs produced cannot be activated. Trojan insertion is prevented as the attacker finds it difficult to find suitable location to place the Trojans because the transition probability of the signals that are being altered by the key gates is unknown to the attacker [12]. Extracting the secret key of the logically locked design has become an extensive effort and it is commonly known as key guessing attacks.

Thus, the main objective of the proposed technique is to build a secured hardware against attacks like brute-force attack, hill climbing attack, and path sensitization attacks.

RLL inserts key gates such as XOR/XNOR gates at arbitrary locations in a circuit. These gates can be configured as buffers or inverters by using the key inputs. If the key gates are left with no further modification in the circuit, then the key bits can be retrieved by inspecting the key gates [11]. Due to fault masking, random insertion of key gates will not propagate the excited faults toward the primary output [1]. RLL is not efficient enough to corrupt a sufficient number of outputs. It also suffers from low Hamming distance and is susceptible to key guessing attacks.

On the other hand, weighted logic locking is inherently immune to the key-sensitization attack. This is because the key inputs are integrated in control gates initially and are not driven directly. The likelihood of any key gate corrupting the design's output is higher in WLL; therefore, fewer key gates can be used to achieve 50% HD, which thereby reduces the execution time drastically.

WLL is also fairly generic which implies that, in order to improve security, it can be combined with other techniques.

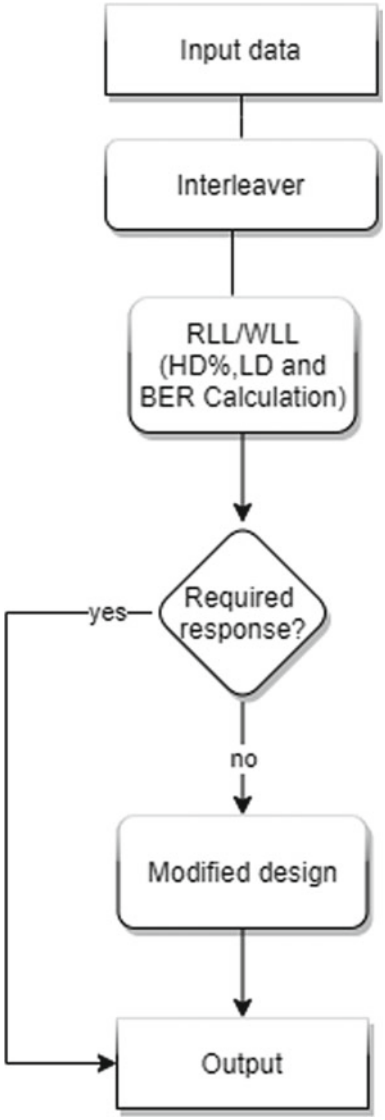



**Fig. 3** Flowchart representation of proposed method

A single key input controlled key gate can be actuated by a random key with a probability of 0.5 (Pact) as the corresponding key bit has a probability of 0.5 to be set as a correct or wrong value. This implies that when a random false key is applied, roughly fifty percent of the key gates have an impact on the circuit. Increasing the Pact of each key gate will result in furthermore actuated key gates when invalid keys are applied, implying that only few of them are needed to achieve the desired HD value.

The Pact can be increased when more than one key input controls the key gate. The key gate is unactuated when right values are given to all key inputs controlling it else, even one wrong key input will actuate the key gate. This can be achieved by using a NAND or AND gate along with the key gates that are XNOR/XOR [16].

Key generation modules need to be embedded into the actual circuit and possess the following:

- Lower area overhead
- Not much increase in power consumption
- Authentication keys to be encoded, to enhance security, which will complicate their true origins and curb logical reverse engineering [9].

In XOR-based LL, functional buffers or inverters are replaced by key gates which introduce uncertainty in retracing the circuit design and the knowing the key values become significant [6].

## 6 Result

In this paper, when logic locking is implemented to secure an interleaved memory, the correct key will give us the correct output as in Fig. 4a and in case of wrong key the wrong output is displayed as in Fig. 4b. The input is loaded after one clock cycle and the output is displayed after 4 clock cycles.

Valid out can be assigned as per the need of the application. The initial clock cycle is used to load the data and hence the output for that cycle remains undefined. Two codes, i.e., two different arrangements of logic locks have been implemented to strengthen the security provided by the logic locking system against external threat.

From the results obtained in case 1 (Fig. 5), it is observed that WLL has much better values (in terms of Hamming distance (HD) and Levenshtein distance (LD) than RLL. But this still did not provide results in the required range of values (50% HD as expected) and this was also confirmed with the calculated BER values. The vagueness of an attacker to the output values, on application of random keys, can be maximized on achieving a 50% HD. Although any percentage change provides ambiguity, 50% or close to 50% is considered desirable.

This prompted for development of the second code (Fig. 6a, b) in which a number of logic locks were increased in general and that increased the overall performance as expected. As shown here, the HD%, BER and LD values obtained are closer to ideal in the case of weighted logic locking (WLL) than compared to that of random logic locking (RLL) for the codes.

So, one thing is observed in common in both the test cases, WLL performs considerably better than RLL and this is seen because in the case of RLL the logic lock is controlled by only 1 key value. In this case, the impact of the wrong key is only 50% (that is because the key can be either 1 or 0 and the probability of any of that being guessed correctly is 0.5). This is why WLL was chosen, in this case the logic lock depends on more than one input, effectively increasing the chances of getting

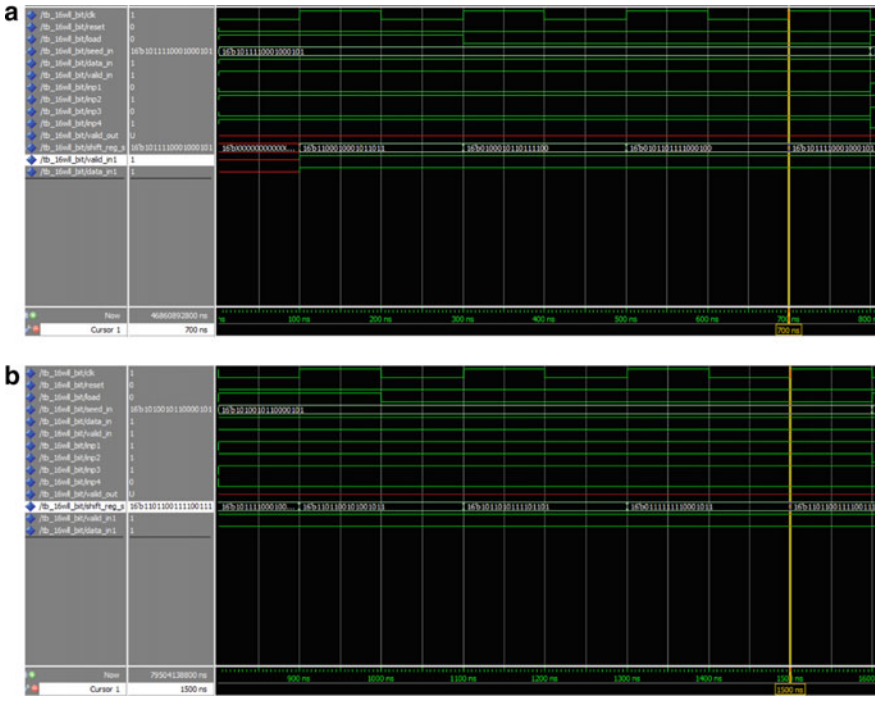



**Fig. 4** **a** Waveform for correct key input for 16-bit data. **b** Waveform for wrong key input for 16-bit data

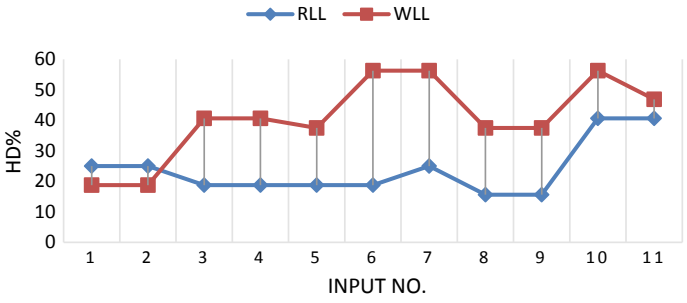



**Fig. 5** Comparison between RLL and WLL HD% obtained for 32-bit data

a wrong result when the wrong key value was entered. Take the case of two keys controlling a logic lock, the chance of the correct key is only ¼ that is 25% but that of the wrong key is ¾ that is 75% which is clearly better than 50% odds. In case of WLL, two or more inputs (apparent key values) are passed as inputs to what is called a control gate. The control gate can be any gate (Fig. 7), (for analysis in this paper, only AND and OR gates are used). The output of the control gate is taken as one of

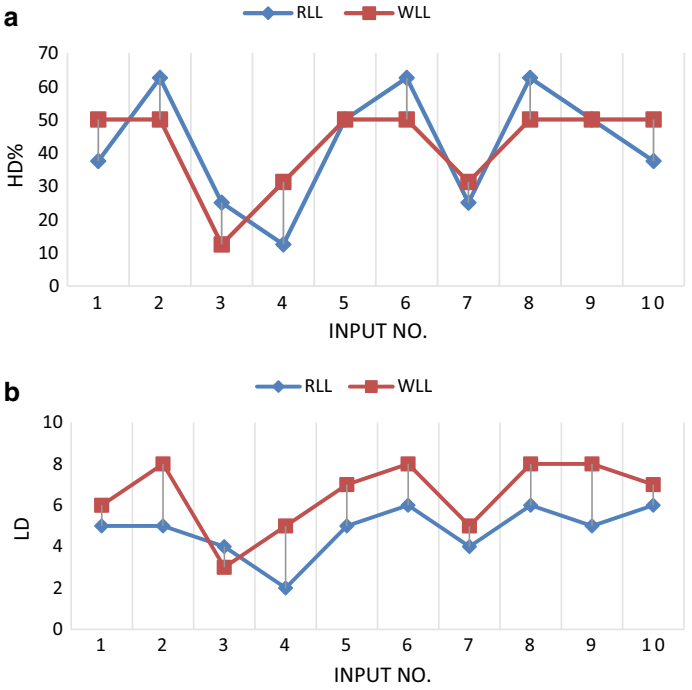



**Fig. 6** **a** HD% calculated for 16-bit data, around 50% for WLL, **b** LD values for both RLL and WLL

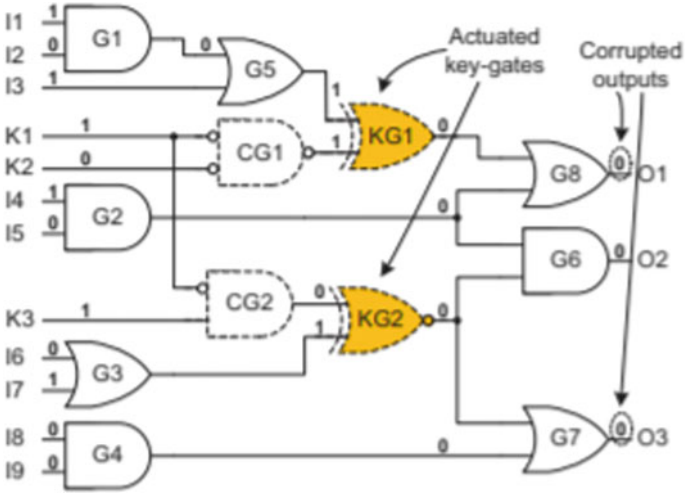



**Fig. 7** An example for the control gate that can be used [16]

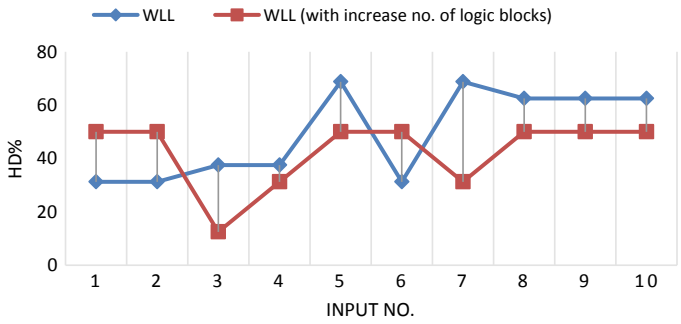



**Fig. 8** Change in the HD% due to addition of more logic locks in the new code

the inputs to the logic lock (either a XOR gate or a XNOR gate), where the other input to the logic lock will be from the circuit the lock is being attached to.

As shown in the graph (Fig. 8), the values obtained can be made by adding logic locks, changing the number of inputs the lock is dependent on, etc. The efficiency of WLL can be further amplified by using increased number of controlling key inputs per key gate; three key inputs shoot up the Pact to 0.88 (7/8), with 4 it extends to 0.94 (15/16), while with 5 it goes up to 0.97 (31/32) [16].

The difference between the two cases obtained has already been explained briefly but to give more insight, in case 1, a section of the circuit was dependent on a set of keys and another section was dependent on another section of keys but this caused a form of discontinuity or reduced the impact of wrong keys on all parts of the circuit so in the second case, the issue was countered by having no particular condition like in case 1. Any wrong key input would impact any part of the circuit thus causing more ambiguity and more anonymity. This makes it more secure from key-sensitization attacks. Figure 9a–c represents the schematic diagram for different bit inputs for the proposed methodology.

## 7 Conclusion

This paper proposes an efficient method for securing data using interleaved memory and logic locking. Two types of combinational logic locking approaches namely random logic locking and weighted logic locking were implemented and the results were compared. Two different codes with varying number of logic locking blocks were simulated for RLL as well as WLL. From the HD and LD values obtained, it was observed that a twofold security method involving WLL on an interleaved memory is an efficient way to establish information security. Interleaved memory accounts for arbitrary arrangement of input data. This memory when further secured by weighted logic locking provides resistance against key-sensitization attacks and promises shorter execution time. WLL provided an optimum 50% HD and acceptable

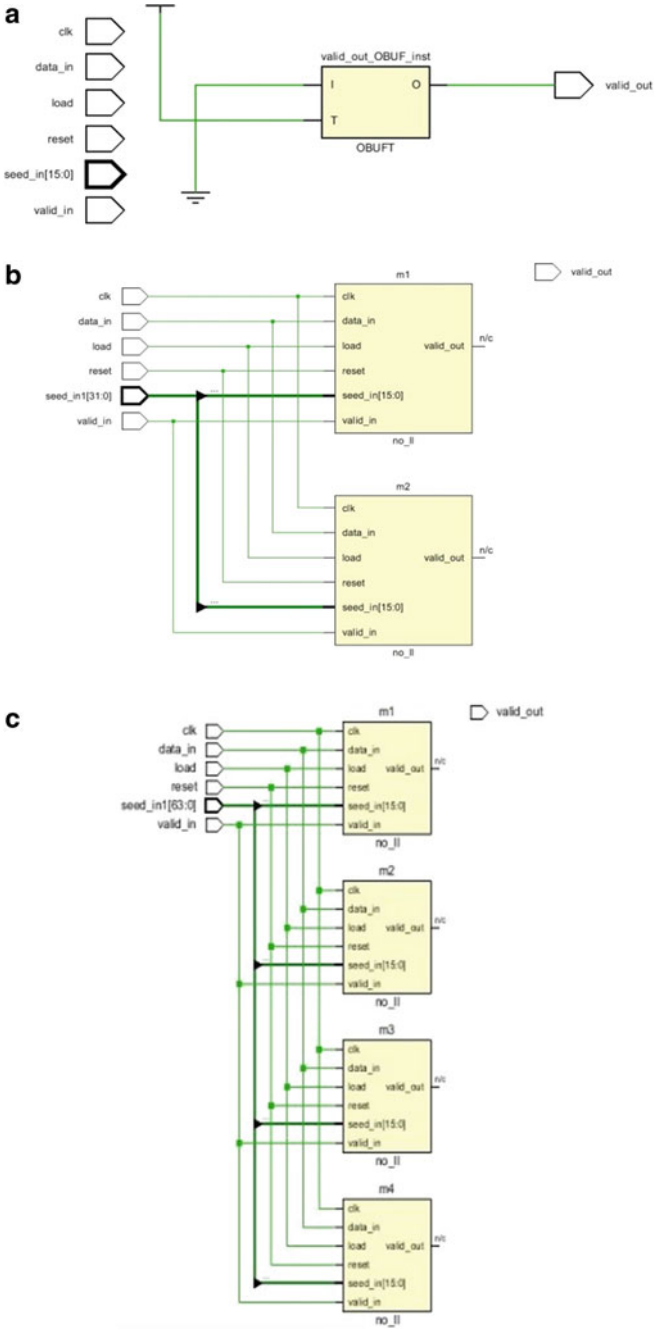



**Fig. 9** **a** Schematic diagram for 16-bit input. **b** Schematic diagram for 32-bit input. **c** Schematic diagram for 64-bit input

LD value for around 50% inputs whereas RLL provides 50% HD to comparatively low number of inputs. Further improvement by adding more logic locking blocks showed an increase in the percentage of inputs giving optimum HD% and LD. Future works can enhance the placement of the logic locking blocks and compare results obtained for each arrangement.

## References


1. Sree Ranjani R, Nirmala Devi M (2018) A novel logical locking technique against key-guessing attacks. In: 2018 8th ınternational symposium on embedded computing and system design (ISED), Cochin, India, pp 178–182. https://doi.org/10.1109/ISED.2018.8704003
2. Raje B, Markam K (2018) Review paper on study of various Interleavers and their significance
3. Torrance R, James D (2009) The state-of-the-art in IC reverse engineering. In: Clavier C, Gaj K (eds) Cryptographic hardware and embedded systems—CHES 2009. CHES 2009. Lecture notes in computer science, vol 5747. Springer, Berlin, Heidelberg
4. Yasin M, Sinanoglu O (2017) Evolution of logic locking. In: 2017 IFIP/IEEE ınternational conference on very large scale ıntegration (VLSI-SoC), Abu Dhabi, pp 1–6. https://doi.org/10.1109/vlsi-soc.2017.8203496
5. Yasin M, Rajendran J, Sinanoglu O (2020) The need for logic locking. In: Trustworthy hardware design: combinational logic locking techniques. Analog circuits and signal processing. Springer, Cham
6. Sengupta A, Mazumdar B, Yasin M, Sinanoglu O (2020) Logic locking with provable security against power analysis attacks. IEEE Trans Comput Aided Des Integr Circ Syst 39(4):766–778
7. Massad MEl, Zhang J, Garg S, Tripunitara MV (2017) Logic locking for secure outsourced chip fabrication: a new attack and provably secure defense mechanism. arXiv preprint arXiv:1703.10187
8. Dupuis S, Ba P, Di Natale G, Flottes M, Rouzeyre B (2014) A novel hardware logic encryption technique for thwarting illegal overproduction and Hardware Trojans. In: 2014 IEEE 20th ınternational on-line testing symposium (IOLTS), Platja d'Aro, Girona, pp 49–54
9. Reddy DM, Akshay KP, Giridhar R, Karan SD, Mohankumar N (2017) BHARKS: built-in hardware authentication using random key sequence. In: 2017 4th ınternational conference on signal processing, computing and control (ISPCC), Solan, pp 200–204. https://doi.org/10.1109/ispcc.2017.8269675
10. Guin U, Huang K, DiMase D, Carulli JM, Tehranipoor M, Makris Y (2014) Counterfeit ıntegrated circuits: a rising threat in the global semiconductor supply chain. Proc IEEE 102(8):1207–1228
11. Yasin M, Rajendran JJ, Sinanoglu O, Karri R (2016) On ımproving the security of logic locking. IEEE Trans Comput Aided Des Integr Circ Syst 35(9):1411–1424
12. Yasin M, Mazumdar B, Rajendran J, Sinanoglu O (2019) Hardware security and trust: logic locking as a design-for-trust solution. In: Elfadel I, Ismail M (eds) The IoT physical layer. Springer, Cham
13. Upadhyaya B, Sanyal S (2009) VHDL modeling of convolutional ınterleaver–deinterleaver for efficient FPGA implementation. Int J Recent Trends Eng 2. LETTERS
14. Roy JA, Koushanfar F, Markov IL (2008) EPIC: ending piracy of ıntegrated circuits. In: 2008 design, automation and test in europe, Munich, pp 1069–1074. https://doi.org/10.1109/DATE.2008.4484823
15. Chen JIZ (2020) Smart security system for suspicious activity detection in volatile areas. J Inform Technol 2(1):64–72
16. Karousos N, Pexaras K, Karybali IG, Kalligeros E (2017) Weighted logic locking: a new approach for IC piracy protection. In: 2017 IEEE 23rd ınternational symposium on on-line

testing and robust system design (IOLTS), Thessaloniki, pp 221–226. https://doi.org/10.1109/IOLTS.2017.8046226
17. Kumar AVP, Bharathi S, Meghana C, Anusha K, Priyatharishini M (2019) Toggle count based logic obfuscation. In: 2019 3rd ınternational conference on electronics, communication and aerospace technology (ICECA), Coimbatore, India, pp 809–814. https://doi.org/10.1109/ICECA.2019.8821935
18. Rekha S, Reshma B, Dilipkumar NP, Crocier AA, Mohankumar N (2020) Logically locked I2C protocol for ımproved security. In: Bindhu V, Chen J, Tavares J (eds) International conference on communication, computing and electronics systems. Lecture notes in electrical engineering, vol 637. Springer, Singapore. https://doi.org/10.1007/978-981-15-2612-1_67

# A Novel IoT Device for Optimizing "Content Personalization Strategy"


**Vijay A. Kanade**

**Abstract** The research paper discloses a theoretical model of a "content personalization device" that functions over an operating system (OS) of any mobile handheld device. The device is in the form of a thin-layered transparent film that fits on the screen of the mobile device. The device operates as a layer over the OS as all the content mediated or allowed by the OS on the device passes through this personalization layer, where it gets filtered. This layer analyzes user interactions with the device and provides filtered content based on the user's response to the content. The externally integrated device tracks the user's device usage pattern and adapts to changing user choices and preferences over time for filtering and displaying relevant content to the user of the mobile device.

**Keywords** Content personalization · Content filtering · User preferences · Feedback loop · User's real-time response

## 1 Introduction

Content personalization refers to rendering Web-based or app-related content to a user based on user's choice or preferences. Web site visitor history is used to provide relevant content to the user in a manner that promotes end-user satisfaction and in most cases helps in lead conversion [1].

Content personalization is generally opted by e-commerce players to double their e-commerce sale. Such personalization is termed as one-to-one marketing as the organization's Web site specifically targets individual customers. Personalization also leads to meeting consumer needs effectively and in an efficient manner, thereby motivating the customer to revisit the Web page in future [2, 3].

However, with the current personalization strategy, one-to-one marketing has reached a point where it has become more burdensome for the consumers rather than of any help. As the customers visit any Web page, they are overloaded with

V. A. Kanade (✉)
Pune, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_47

personalized ads, social media feeds that are essentially taxing. These ads are sometimes not just irrelevant for the user, but they also tend to gulp a lot of Internet data unnecessarily [4].

Further, personalization is needed in situations where apps tend to send out irrelevant notifications that the user may not be interested in at a particular time. Here again, these notifications are more disturbing similar to the above ads, rather than of any value. Hence, there seems to be a long pressing need to identify a solution for content personalization that is not just specific to the user but takes into consideration user behavior, device usage pattern, and user context before delivering any content. The research paper provides a theoretical model of a novel IoT device that addresses this problem of content personalization. The IoT device also puts a break on excessive personalization opted by most marketing agencies [5].

## 2 Content Filtering Techniques

Content filtering is employed strategically by various business verticals. Some of these techniques are as listed below [3]:

1. Collaborative content filtering

Here, data from different Web sites is collaborated and filtered to provide user-relevant data that may give better user experience (e.g., e-commerce) to the user.

2. User-specific profiling

Creating personalized Web page for a user based on user-specific data collected from various sites.

3. Analytic tools

Data analysis tools are used to predict likely future interactions of the user based on the user's activity on the site.

### *2.1 Real-Life Examples of Content Personalization*

Amazon, Netflix, and YouTube are some of the common platforms that disclose content personalization for better user experience [6, 7].

1. Amazon generates product recommendations based on page views, user purchase history, and user behavioral data.
2. Netflix recommendation engine uses user's viewing history, interactions, content title info (i.e., genre, cast, etc.), duration of sessions, device type, and time of the day.

3. YouTube provides recommendations based on the user's past history, interaction patterns, and from demographic data where content consumed by various users with similar personas is recommended.

However, these recommendations do not take into consideration the user context or the content type that is being consumed in real time. This leads to user distraction many a times. For example, consider YouTube recommendation engine having one-to-one marketing ads. YouTube ads are difficult to bare for consumers, as they seem to pop-up after every "n minutes." Now, when the user is amidst a serious video content, such interruption by ads becomes disturbing and forces the consumer to view the ad. Although personalization enhances customer experiences, however, its biggest challenge today is that it should not act as a source of distraction to its customers [8].

## 3 Novel IoT Device

### *3.1 Hardware*

The proposed novel IoT device is a thin transparent sheet that fits on the top screen of the mobile device. The arrangement of the device is similar to that of a screen guard that is adopted by most mobile phone users for protection against screen damage. The device is provided with a micro-USB cable that is used to plug-in and integrate it with the mobile device (Fig. 1).

### *3.2 Architecture*

The IoT device consists of a secure file system that sits on top of the original OS of the electronic device. On integration of the IoT device with the mobile device (i.e., by plugging-in the micro-USB cable), it synchronizes with all the repositories of operating system(s). This may include AppStore for MacOS for Iphone-based system, Google Play Store for Android-based system, Windows for Microsoft. The secure file system is provided with a monitoring module through which the data flow of the device takes place. This module monitors and tracks the inter-application data traffic occurring within the device as well as the data communicated by the device via servers when using the Internet data. Therefore, user interaction with the device is closely monitored by the secure file system of the IoT device.

The architecture of the IoT device is shown in Fig. 2.

**Description**

In Fig. 2, the secure file system acts an independent layer above the OS. Such an arrangement ensures that the device data exchange between various apps (AppN) or Web sites (WebsiteN) is closely tracked leading to a better understanding of the user

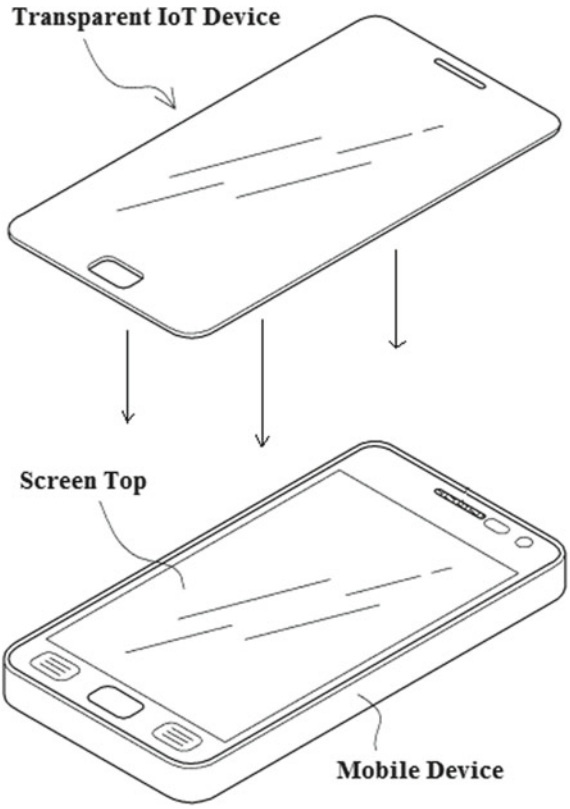



**Fig. 1** Theoretical model of transparent IoT device. (Micro-USB cable not shown)

interactions with time. The secure file system helps to gauge user context based on the feedback (i.e., user response) that the user provides for the live content that is currently being played on the mobile device. This allows better content filtering as it takes into account the changing user's taste over time as content is being delivered to the user.

### *3.3 Data Flow*

The general data flow of proposed theoretical model is shown in the below flowchart (Fig. 3).

The steps involved in the operation of the novel IoT device are summarized below:

Step 1: Monitoring data exchange
Step 2: Verify user context [interaction pattern, behavioral data]
Feedback loop; [user response on live content]
Step 3: Allow content delivery or block content

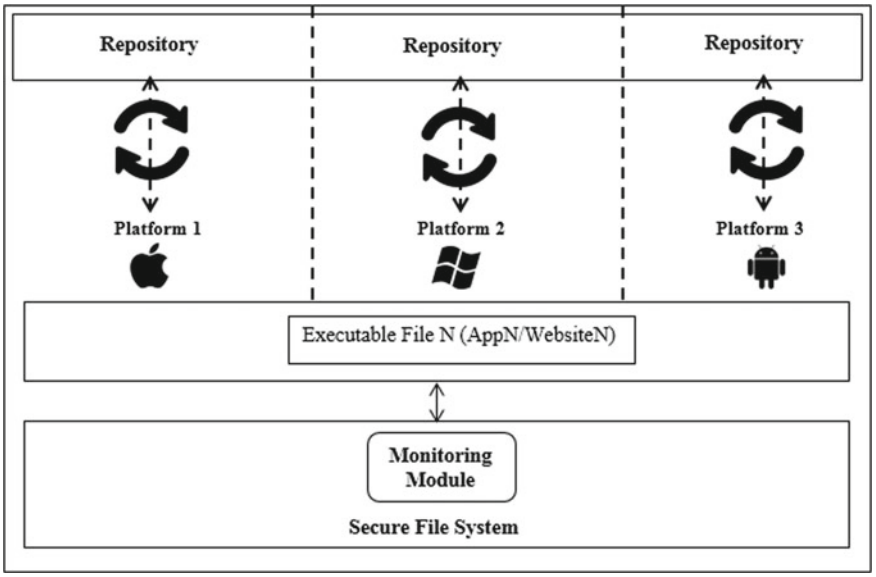



**Fig. 2** Theoretical architecture of IoT device

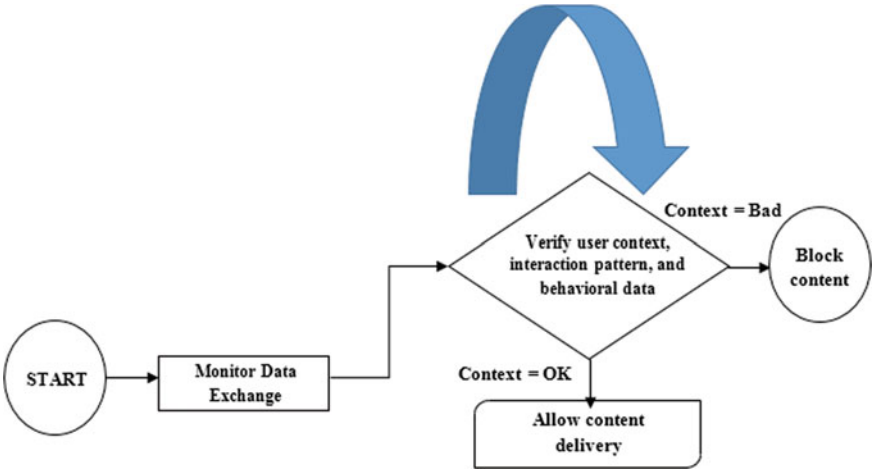



**Fig. 3** Data flow during content personalization

Context value; [good or bad]
Repeat; [loop-in > Step 1].

## 4 Preliminary Research Observations

Currently, the end users are bombarded with excess personalized information, where the marketer is missing on the actionable content that would derive insights needed to drive user engagement and earn substantial revenue.

A study was recently conducted to discover thoughts and feelings of the college students regarding the annoying marketing tactics adopted by most marketers today [9]. About 273 students from a large public university located in the southern USA participated in the study. The insights derived from the study are as tabulated in Table 1.

Hence, the proposed theoretical model of the IoT device is designed to lower this impact of excessive personalization on customers by controlling the type of content being delivered to the end users based on user's real-time response to the provided content, i.e., feedback.

## 5 Applications

The proposed research is specifically useful in following scenarios:

1. Notifications and ads are delivered to the user by tracking user feedback for certain content type that is already being delivered.
2. Consider a scenario, where the calendar shows meeting scheduled at 12–1 pm. In such cases, notifications are not provided to the user, as the phone can be put on DND mode. However, if something very critical pops-up while the DND mode is on, then the proposed IoT device will monitor this content and flag it, so that the content is delivered to the user irrespective of the DND mode. This can happen for SOS emergency calls, urgent meeting emails, or in cases when someone sends a serious text message over WhatsApp or any social media.
3. Another case can be considered of Apple IPhone, where at night the phone enables DND, so that all the notifications are automatically silenced until user's wake-up time. However, some important texts, updates, or notifications can be missed due to this facility. Hence, the disclosed IoT device checks the semantics of the data being communicated to the mobile device and thereby decides if the content should get through the DND mode. An example of this could be, allowing phone calls from contacts saved as Mom, Dad, Family, etc., to get through, but denying access to rarely communicated contacts.
4. Consider another scenario where the user is continuously skipping ads on YouTube. In such a scenario, as the user usage pattern is already being tracked by the IoT device, next time YouTube decides to play an ad, that content will be automatically skipped or filtered, thereby avoiding its delivery to the user. Further, the discussed marketing ads also consume excess Internet data, which can be avoided by using the disclosed IoT device.

**Table 1** Study responses [9]

| Annoying marketing tactics | Subcategories | Marketing quotes | Percent |
|---|---|---|---|
| Aggressive tactics | | "Marketing that keeps trying to influence a customer to buy a certain product even when they already turned the product down" | 41 |
| Repetition | | "When companies use the same commercials over and over" | 23 |
| Obnoxious tactics | | | 15 |
| | Garish/loud ads | "Any marketing that is loud, persistent, or vulgar, or things that are spelled incorrectly" | 8 |
| | Irritating | "Marketing that is irritating to any of the 5 senses of a target customer" | 4 |
| | Distracting | "Ads before YouTube videos, Pandora, etc. Any ad that keeps you from doing what you are trying to do" | 1 |
| | Low quality | "Marketing that does not meet today's standards" | ~1 |
| Irrelevant content | | | 7.3 |
| | Nonsensical | "Ads that seem to have no plan" | 4 |
| | Uninformative | "Marketing that tells nothing about the product. It just tries to be funny or flashy to sell me something" | 2 |
| | Product focused | "Only concentrating on selling the product, not caring about the customer" | 1 |
| Tactics which cause avoidance | | "Bugging a customer to the point that they want nothing to do with your product" | 5 |
| Consumer discomfort | | | 32 |
| Violates social norms | | | 14 |
| Out of the ordinary tactics (creates weird feeling) | | "Marketing that uses techniques that we think are odd or unusual to get our attention" | 13 |

## 6 Conclusion

The proposed research discloses a theoretical model of a novel IoT device to fine-tune and optimize the content personalization strategy. The personalization mechanism takes into consideration user context, user behavioral data, and user interaction for filtering the content. The proposed research removes unwanted content delivered via a modern one-to-one marketing strategy and recommendation engine. This gives the

user better control over the Internet data as only user-relevant content is allowed to pass through the IoT device.

**Acknowledgements** I would like to extend my sincere gratitude to Dr. A. S. Kanade for his relentless support during my research work.

## References

1. Vesanen J (2007) What is personalization? A conceptual framework. Eur J Mark
2. Shen A (2014) Recommendations as personalized marketing: insights from customer experiences. J Serv Mark
3. Rouse M Content personalization
4. The curious case of irrelevant marketing, Automat
5. Gerard M (2020) Overcoming common personalization problems: data quality and scalability, 25 August 2020
6. Arora S (2016) Recommendation engines: how amazon and Netflix are winning the personalization battle, 28 June 2016
7. Kumar A (2020) YouTube's recommendation engine: explained, 30 January 2020
8. Choi J (2019) 3 challenges of personalization at scale, 1 May 2019
9. Moore RS, Moore ML, Shanahan KJ, Horky A, Mack B (2015) Creepy marketing: three dimensions of perceived excessive online privacy violation. Mark Manag J. Spring

# IoT Based Self-Navigation Assistance for Visually Impaired


**Nilesh Dubey, Gaurang Patel, Amit Nayak, and Amit Ganatra**

**Abstract** One of the major challenges faced by visually challenged people is self-navigation in unknown environments. They often tend to get hurt by objects that they cannot feel using their hands or a walking cane, as certain objects are hard for a blind person to detect by just using tapping their walking cane. To avoid obstacles and navigate through a new environment, a smart belt for the visually impaired is performed in which he/she can continuously detect the obstacles around the user with its sensors that span the entire 360° of his/her field of view. Whenever an obstacle is in a nearby range, sufficiently enough to cause a hindrance, the device will give sensory cues to the user about the location of the obstacle and family members are also tracked them using GSM and GPRM system.

**Keywords** GSM · Obstacle detection · Visually impaired · HC-05 sensor system

N. Dubey (✉) · G. Patel
Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research, Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology, Changa, India
e-mail: nileshdubey.ce@charusat.ac.in

G. Patel
e-mail: gaurangpatel.dcs@charusat.ac.in

A. Nayak
Department of Information Technology, Devang Patel Institute of Advance Technology and Research, Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology, Changa, India
e-mail: amitnayak.it@charusat.ac.in

A. Ganatra
Department of Computer Engineering, Charotar University of Science and Technology, Changa, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_48

# 1 Introduction

Unlike us, blind or visually challenged people are not blessed with the same gifts, and consequently, they cannot perform the same tasks with the same parity as what a normal individual could do. In an entirely novel place, they might be able to dodge the obstacles coming in their way with the help of the walking cane but knowing where to go seems like an uphill climb without any external help as they are unfamiliar with the layout. This could lead to getting lost sometimes, which would be a difficult situation to get out of. The current solution also renders them with only one usable hand while walking, which in some cases may be a minor inconvenience. Countless other problems are not mentioned here but affect them daily.

The project aims to mitigate the most common problems of their struggle and make their lives comparable to ours. It takes advantage of a simple wearable belt that everyone can adorn. On the belt, there would be ultrasonic SR-04 sensors fitted, about six of them which handle their own respective direction. Four sensors are entirely dedicated to the $X$-$Y$ plane, i.e., the horizontal plane located at about the waist height of the wearer. They continuously pulse in the four major directions, i.e., North, East, West, and South. This way, the entire 360° field of view of the blind person is covered by the sensors. The extra two sensors cover the vertical plane, and they check for any obstacles on the floor which may come in the way of the person. Hindrances at the head height of the person should also be taken care of, which is done by the sensor aligned at the pre-decided angle with the horizontal plane. For the additional features, the distress alert feature helps them to stay connected wherever they are. In any event of trouble, this feature will notify all the people which were set earlier when setting up the device. Additionally, with the help of an Android app, connected to the device via Bluetooth, it can help in navigating walking distances by giving turn by turn directions to the user via the handset connected by the auxiliary port with the Android phone. Through this app, blind people walk safely outdoors or in traffic areas.

# 2 Existing Work

## *2.1 Smart Mobility Aid for the Visually Impaired Society*

Electronics is a domain that is constantly growing and developing [1–4]. There are approximately 15 million people who are blind in India. Major developments square measure worn out a sensible for seeking a well and smarter life and welfare toward the blind society. This paper [4–6] recommends and pursues a thought of removing the stick and mounts these sensors on the visually handicapped individual body itself.

### 2.2 Real-time Dangling Objects Sensing

The initial style of mobile ancillary device for pictorial impaired [7–10] was discussed. This analysis planned a mobile real-time dangling objects sensing (RDOS) model that is mounted on a front of a cap for obstacle detection. This device uses relatively low-cost audible sensing parts which provide balance sense for blinds to know the anterior projected things. The RDOS gadget can the executives the sensor's front point that is at the client's body stature and upgrades the detecting precision. The RDOS gadget can manage the sensor's front angle that is at the user's body height and enhances the sensing exactness [11]. Two major required algorithms are to live the height-angle activity and un-hearable detector alignment and planned unit space. The analysis team to boot combined the RDOS device with mobile automation devices by act and Bluetooth to record the walking route.

### 2.3 Assistive Infrared Sensor-Based Smart Stick for Blind People

In this article, the authors proposed a lightweight smart stick that is relatively inexpensive, very user-friendly, and easy to use [12–14]. The smart stick gives a quick response to the user about the obstacles ahead and helps them with better mobility and uses low power consumption. The stick utilizes infrared technology; a cluster of infrared sensors can observe stair-cases and totally different obstacles presence at intervals the user path. As a stick could be an acquainted object to a blind man thus employing a good stick for higher quality and obtaining the device would be terribly straightforward [15, 16]. The experimental results provide smart accuracy and thus the stick is prepared to watch all of the obstacles [17].

## 3 System Components

### 3.1 Arduino Uno

The ATmega328 AVR microcontroller is the core of the Arduino Uno R3 microcontroller board. It consists of 20 digital input and output pins which comprise six PWM outputs and six analog inputs. A computer program named Arduino Uno IDE is used to load the code from the computer to the Arduino. The Arduino has widely used in the world which makes it an extremely simple approach to begin working with embedded electronics.

### 3.2 Ultrasonic Sensor SR-04

The ultrasonic sensor SR-04 is used to measure the distance between the sensor and some object. The way distance is measured by the sensor is that it emits a sound, usually about 40 kHz which falls in the inaudible range for humans, but the sensor is capable of picking it up. After emitting the sound, the sensor waits for the sound to bounce back from the object, and it measures the time taken. With the help of time taken, it gets conceivable to calculate the distance between the object and the sensor. The general formula speed is equal to distance divided by time is used here. It is rearranged to be distance which is equal to speed multiplied by time. The next step is to simply plug in the values for the speed of sound and the time measured by the sensors in the previous step and obtains the distance. Here, the distance calculated would be the round-trip time, which means the distance from the sensor to the object and the distance between the object and the sensor. To solve this, just divide the distance by two.

$$\text{distance} = \frac{\text{speed of sound} * \text{time taken by sound}}{2} \tag{1}$$

### 3.3 Bluetooth Module

HC-05 module is an unbelievably essential Bluetooth SPP (Serial Port Protocol) module, proposed for direct far off successive affiliation plan. The HC-05 Bluetooth module can be used in a master or slave design, making it a viable answer for remote correspondence. This sequential port Bluetooth module is a totally qualified Bluetooth V2.0+EDR (enhanced data rate) 3Mbps modulation, and it contains an absolute 2.4 GHz radio handset and baseband. It uses CSR Blue center 04-External single-chip Bluetooth structure with CMOS development and with adaptive frequency hopping feature (AFH). Bluetooth, a ton of explicitly Bluetooth low energy (BLE), has become the overwhelming innovation for associating Internet of Things (IoT). The innovation is also ready to keep up the quality correspondence range all through the trade. This low energy form of Bluetooth may totally affect IoT innovation, by giving gadgets the adaptability to existing and with progress perform during a huge decision of use situations. Bluetooth empowers low force correspondence between gadgets that are in the closed closeness of each other existing instrument.

## 4 System Design

This is the basic format of how the sensors are placed around the belt and what field of views will they cover. The lines *a* and *b* work on the XY plane, toward the front

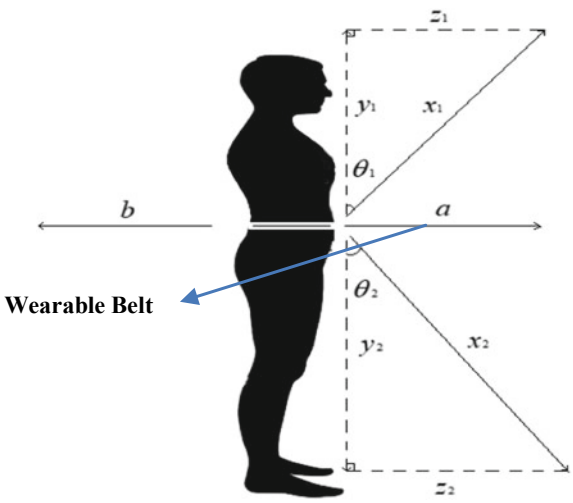



**Fig. 1** Representation of the proposed system with wearable belt

and rear of the wearer, respectively. Both sensors are not person dependent, and they will have the same configuration regardless of the wearer because they work in a straight line perpendicular to the user. The height of the wearer will not affect their function in any way (Fig. 1).

The next two sensors sit at a little distance from the sensor with the line of sight *a*. The problem with fitting all the sensors equidistant to one another is that two sensors will fall on either side of the torso which is the normal resting position for both the arms. Hence, putting them on the sides will give erroneous results as hands constantly interfere with the reading of the sensors. The remaining two sensors have a slightly tricky implementation. Their implementation depends on the height of the wearer as well as the waist height of the wearer concerning the ground. The sensors are placed at an angle over and below the front sensor. The bottom sensor, with the line of sight $x_2$ detects the steps or little obstacles along the way of the user. The problem with placing the sensor pointing straight down is that due to the walking it is impossible to detect the distance accurately. Also, it is pointless to do that as the detection of the obstacles takes longer time. Hence, the angle $\theta_2$ is placed concerning the vertical plane. The distance $x_2$ will depend on the height $y_2$ at which the user wears the belt on his waist as the Pythagoras theorem will be applied here.

$$x_1^2 = y_1^2 + z_1^2 \tag{2}$$

Also, to calculate without the distance, the formula is very useful.

$$\cos\theta_1 = \frac{y_1}{x_1} \tag{3}$$

The sensor which detects above the head of the user works in a very similar way with a slight difference. The distance $x_1$ is decided on $y_1$, but $y_1$ is not the height

**Table 1** Values are taken for the respective variable and generated values of $\theta$

| Parameters | $x_1$ | $y_1$ | $z_1$ | $\theta_1$ | $x_2$ | $y_2$ | $z_2$ | $\theta_2$ |
|---|---|---|---|---|---|---|---|---|
| Data | 128.6 | 91 | 91 | 45° | 142.8 | 101 | 101 | 45° |

of the wearer; it represents a distance that is slightly above the height of the wearer from the waist. This is done because objects near the head of the user can pose a significant threat while walking. The code is designed in such a way that it will trigger the buzzer if anything is detected closer than 20 cm to the height of the head of the user.

With the current tested data, the following table is prepared using the device (Table 1).

### *4.1 Design Requirements*

(1) It has one input because of having one degree of freedom.
(2) It is a coplanar mechanism with at least four links.
(3) It has at least one combine pair to form a non-uniform output speed.
(4) It has at least one gear pair to vary the uniform resultant speed.
(5) It has a ground link to support or constrain other links.
(6) It has one output link.

### *4.2 Design Constraints*

(1) The number of links should be four or five links.
(2) The frame must be a link with three joints or more to have a firm support.
(3) At least one cam pair must be incident to the frame.
(4) The input link must be adjacent to the frame with a revolute joint.
(5) The input link, the output link, and also the frame should be appointed on completely different links.

## 5 Experimental Results

After rigorous testing, it can be concluded that six sensors are the optimal solution required for this device to work efficiently. It creates a good balance between the working and affordability of the device, and it provides enough sensory cues for the visually impaired person. Anymore several sensors will confuse the user and exacerbate the problem instead of alleviating it. To implement the sensors in many different combinations of wearables, environments found that a belt can be used in

**Table 2** Classification of hurdle based on sensor readings

| Type of hurdle | Type of alert |
|---|---|
| Non-uniform path | Voice alert Beep 1 |
| Small object | Voice alert Beep 2 |
| Large object | Voice alert Beep 3 |

all kinds of weather and can be worn by almost everyone. Also using the HC-05 sensor instead of the SIM9000 sensor proves to be a wise choice as the recipient of the SIM9000 sensor can be poor at times which prevents the distress feature from working. With the connection of an Android phone via a Bluetooth sensor to the device, accurate latitude and longitude can be sent with precision (Table 2).

## 6 Conclusion

The project work aims to solve an underlying problem that has been overlooked for years. The technology for the masses has seen an exponential growth, while it has been stagnant for a certain demographic. This device will replace a tried and tested practice used by blind people, i.e., walking and navigating with the help of canes. The benefits that this product will provide to the users are subtle yet revolutionary. The simple convenience of having both hands-free while still being able to walk freely in a known or unknown location should be available to everyone. This device helps in making blind people more independent with the use of SOS features and more familiar with technology in general. Thus, the obstacles were detected using the SR-04 sensors via measuring their latency of the 40 kHz sound signal and processing using the algorithm used by the Arduino Uno.

## References

1. Sailaja P (2020) IoT and ML based periodic table for visually impaired. J Adv Res Dyn Control Syst 12(SP7):2673–2682. https://doi.org/10.5373/jardcs/v12sp7/20202404
2. (2020) An IoT based smart assistive device for the visually impaired. In: 2020 IEEE region 10 symposium (TENSYMP). https://doi.org/10.1109/tensymp50017.2020.9231026
3. Choudhary S, Bhatia V, Ramkumar K (2020) IoT based navigation system for visually impaired people. In: 2020 8th international conference on reliability, Infocom technologies and optimization (Trends and Future Directions) (ICRITO). https://doi.org/10.1109/icrito48877.2020.9197857
4. Mudaliar MD, Sivakumar N (2020) IoT based real time energy monitoring system using Raspberry Pi. Elsevier
5. (2019) Assistance for visually impaired people based on ultrasonic navigation. Int J Innov Technol Explor Eng 8(11):3716–3720. https://doi.org/10.35940/ijitee.j9887.0981119
6. Shanmugam M, Victor J, Gupta M, Saravana Kumar K (2017) Smart stick for blind people. Int J Trend Res Dev (IJTRD). ISSN 2394-9333

7. Sourab BS, Sourab BS, D'Souza S Design and implementation of mobility aid for blind people, India. ISSN 978-1-4799-8371-1/2015 IEEE
8. Nada AA, Fakhr MA, Seddik AF In: Assistive infrared sensor based smart stick for blind people. In: IEEE conference, Egypt
9. Lin CH, Cheng PH, Shen ST Real-time dangling objects sensing: a preliminary design of mobile headset ancillary device for visual impaired. ISSN 978-1-4244-7929-0/14/2014 IEEE
10. EL-Koka A, Hwang G-H, Kang D-K Advanced electronics based smart mobility aid for the visually impaired society, Korea. ISSN 978-89-5519-163-9
11. Siddesh GM, Manjunath S, Srinivasa KG (2016) Application for assisting mobility for the visually impaired using IoT infrastructure. In: 2016 international conference on computing, communication and automation (ICCCA), Noida, pp 1244–1249. https://doi.org/10.1109/ccaa.2016.7813907
12. Mala NS, Thushara SS, Subbiah S (2017) Navigation gadget for visually impaired based on IoT. In: 2017 2nd international conference on computing and communications technologies (ICCCT), Chennai, pp 334–338. https://doi.org/10.1109/iccct2.2017.7972298
13. Shreyas K, Deepak G, Ramaiah NS IoT based route assistance for visually challenged, 28 March 2019. Available at SSRN https://ssrn.com/abstract=3361536 or http://dx.doi.org/10.2139/ssrn.3361536
14. Shaha A, Rewari S, Gunasekharan S (2018) SWSVIP-smart walking stick for the visually impaired people using low latency communication. In: 2018 international conference on smart city and emerging technology (ICSCET), pp 1–5
15. Verma P, Sood SK (2017) Cloud-centric IoT based disease diagnosis healthcare framework. J Ambient Intell Humanized Comput 9(5):1293–1309
16. Velázquez R, Pissaloux E (2017) Constructing tactile languages for situational awareness assistance of visually impaired people. Mob Vis Impaired People 597–616. https://doi.org/10.1007/978-3-319-54446-5_19
17. Kunta V, Tuniki C, Sairam U (2020) Multi-functional blind stick for visually impaired people. In: 2020 5th international conference on communication and electronics systems (ICCES), Coimbatore, India, pp 895–899. https://doi.org/10.1109/icces48766.2020.9137870

# An Overview of Cyber-Security Issues in Smart Grid


**Mayank Srivastava**

**Abstract** A smart grid is an updated version of the traditional electrical grid that uses Information and Communication Technology (ICT) in an automated fashion for the production and distribution of electricity. Several attributes like efficiency, sustainability, and reliability are improved in the smart grid as compared to the traditional grid. Smart grid gives significant benefits for the entire community, but their dependence on computer networks make them vulnerable to various kinds of malicious attacks. This article focuses on identifying the various cyber-security issues of different areas of the smart grid which are prone to vulnerabilities. Finally, the possible solutions for resolving the cyber-security issues in the identified areas for making the smart grid more secure were analyzed.

**Keywords** Smart grid · ICT · Smart devices · Cyber-security · Vulnerability

## 1 Introduction

Smart grid technology offers many benefits to the entire society [1]. However, the integration of computer networks into the smart grid makes society vulnerable to different types of malicious attacks. Also, such integration makes the way for the privacy issues of the customer. The ICT is the basis of the underlying smart grid infrastructure, which is required for the successful operation of the smart grid. Henceforth, it becomes necessary to address the security issues of ICT in this domain. Finally, one can say that it is important to address the cybersecurity issues in various phases of the smart grid [2].

Besides this, the unintentional compromises of the network infrastructure due to user-oriented errors, apparatus failures, and natural blows also need to be considered. Securing the entire facility of the smart grid is a daunting task [3]. The vulnerabilities of the existing technology infrastructure must first be identified and analyzed, before applying the management process to mitigating them. But there are certain barriers

M. Srivastava (✉)
Department of CEA, GLA University, Mathura 281406, UP, India
e-mail: mayank.srivastava@gla.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_49

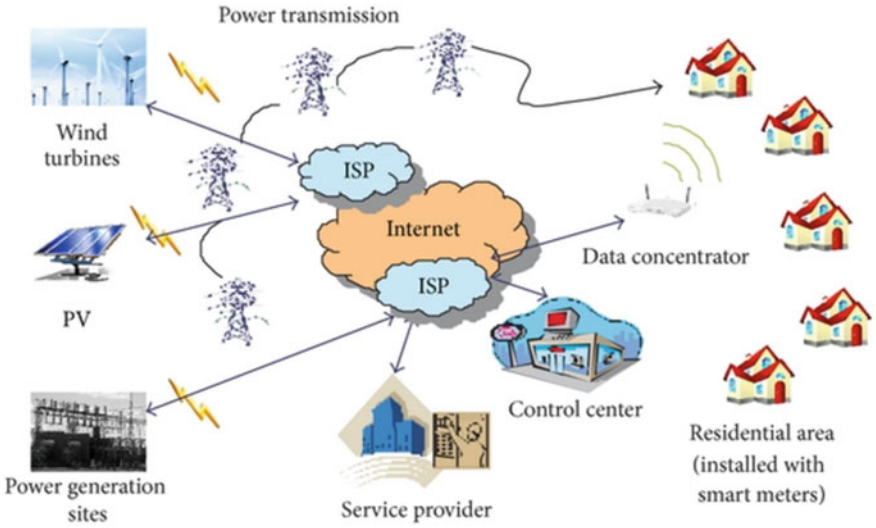



**Fig. 1** Smart grid architecture [5]

such as organizational issues, lack of technical skills, and awareness concerns that prevent us to achieve this objective. [4].

A smart grid basically incorporates a computer network with the existing grid infrastructure to analyze and distribute data related to energy consumption. In the smart grid, the energy producers and energy consumers interact intelligently which eventually serves the purpose of saving a lot of energy. In Fig. 1, the different forms of energy generation schemes like wind turbines, photo-voltaic (PV), and power generation sites are used for power generation. The generated power is distributed to various categories of customers. The figure also shows the interconnection of various nodes across the entire transmission network of the smart grid. The various network components used in the infrastructure are data concentrator, control center, Internet Service Provider (ISP), and various sensors. The sensors are used to give accurate information about different aspects of smart grid architecture.

## 2 Review of Literature

Different researchers are working on various domains of the smart grid. Some of the major contributions of the research done in this area are as follows.

Chebbo [3] focuses on the importance of smart technologies, smart vision, smart processes, and finally smart stakeholders for energy innovation. Tuballa et al. [4] present an outline of the associated technologies of the smart grid. The paper also mentions the importance of the latest technologies in influencing the existing grid. Mrabet et al. [5] provide a deep understanding of the security vulnerabilities and their

possible solutions. The paper gives a cyber-security strategy to deal with various kinds of cyber-attacks. Wang et al. [6] put a thought on the hybrid structure of the computer networks that connect the energy providers and consumers in a smart grid. The author then emphasizes the challenging problems of network reliability and security. Bigerna et al. [7] perform the literature review about the social costs affecting the development of the smart grid. The paper finally presents the opportunities and challenges in the business, applications, policy, and security issues of the smart grid.

Kappagantu et al. [8] specify that the three main objectives of cybersecurity that need to be addressed are availability, integrity, and confidentiality. It also emphasized on multilayer structure of the smart grid, where each layer is having specific security concerns. Finally, it states that the use of advanced techniques is essential to tackle sophisticated cyber threats. Baumeister [9] proposes a study that is divided into five different smart grid security categories namely process control, communication protocol, smart meter, simulation for security analysis, and power system state estimation for achieving smart grid security. Wang et al. [10] present a comprehensive review of different cybersecurity issues for the smart grid which focuses on various vulnerabilities and solutions concerning the smart grid.

Yang et al. [11] discuss the various cyber-attacks and their possible solutions which are crucial for the expected operation of the smart grid. The paper also presented its insight on the critical aspects related to cyber-security of smart grid-like interdependency, vulnerability, etc. Pearson [12] presented a study on the use of ICT in Europe's electricity sector. The article highlights that increased reliance on ICT in the electricity sector will open up new vulnerabilities that will undermine the advantage of the smart grid. It also explains that the European Union (EU) has to mitigate these challenges to avoid a possibly expensive technical lock-in. Yan et al. [13] summarize the different cybersecurity issues and vulnerabilities related to the smart grid. It also presents a survey that focuses on the current solutions to different cybersecurity issues of smart grids.

Aillerie et al. [14] present the report which identifies significant issues in cybersecurity policy design for the International Smart Grid Action Network (ISGAN) membership. The paper discusses the smart grid architecture along with fast-changing cyber threats. The paper also proposes certain hardware designs for improving the security of the smart grid. Berthier et al. [15] focus on using secure protocols to prevent the network from being exploited. It also emphasizes the importance of intrusion detection system (IDS). Sou et al. [16] explore the smart grid cybersecurity problem by analyzing the false data attacks-based vulnerabilities of electric power networks. Here, the analysis of the problem is based on the constraint cardinality minimization problem. It was shown in the paper that a relaxation technique provides an exact optimal solution to the cardinality minimization problem.

## 3 State Estimation of Smart Grid

Figure 2 shows another basic infrastructure of the smart grid. It can be seen from the given figure that it includes cloud-based servers, power transmission lines, smart grid operators, and transmission lines to the industrial, commercial, and residential customers. The whole smart grid infrastructure is based on smarter networks, which are mainly used for transmission and distribution purposes [17]. The smart grid makes use of mostly new technologies to improve the transmission as compared to the traditional system. Advanced distribution automation (ADA) technologies along with advanced metering infrastructures (AMI) provides the required intelligence to the power grid infrastructure to meet out its defined expectations.

The smart grid acts intelligently to integrate distributed energy generation from different energy sources. A smart grid helps mainly in conserving energy, increasing reliability and transparency, reducing cost, and making the entire process of energy generation and distribution more efficient [18]. In a smart grid, the data between consumers and the grid operator is exchanged by using secure communication channels based on encryption. Here, the homomorphic data encryption techniques are used to provide data privacy over the cloud.

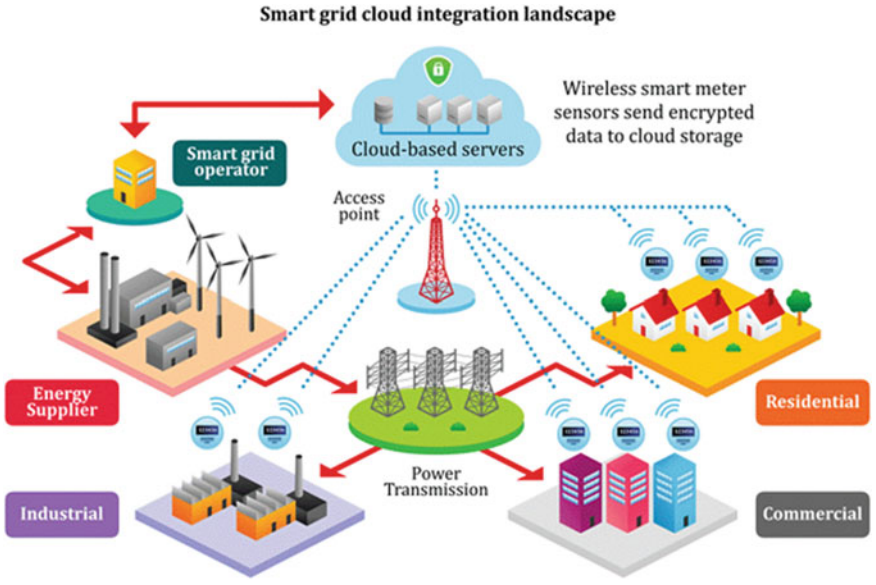



**Fig. 2** The smart grid architecture [17]

## 4 Cyber-Security Issues in Smart Grid

Smart grid infrastructure dominantly uses smart devices that are prone to vulnerabilities. As the smart grid makes heavy use of ICT, this gave the attackers a possibility to exploit the different vulnerabilities of ICT to disrupt the normal operation of the grid. A smart grid is an interconnection of heterogeneous systems, where each system may be having its own set of technologies and communication equipment. This heterogeneity, diversity, and complexity make the security of the smart grid highly difficult. The objective of this section is to highlight the major cyber-security issues of different areas of the smart grid, which is having vulnerabilities and can be exploited by the attackers to do malicious activity [19]. The various cyber-security issues in different areas are as follows:

1. **Architecture**: The weakness of the smart grid architecture can be exploited by the intruder for a possible attack. For example, vulnerabilities related to the operating system, server, etc. can be exploited to perform malfunctioning of the system.
2. **Communication protocols**: Devices can be compromised if they are communicating over an insecure line. Some of the widely used wireless protocols like Zigbee, Wimax, Wifi are also having vulnerabilities.
3. **Interfaces**: Web-based smart grid applications are subject to several vulnerabilities. The vulnerabilities related to the application can be exploited to make the system error-prone.
4. **Home area networks (HAN)**: Vulnerability can also exist in smart equipments within HAN. Network parameters of HAN can be identified to launch various network attacks.
5. **Human factor**: Social engineering techniques can be adopted by the attacker to access customer accounts and to change their settings. For example, by using phishing attacks, the attacker can get the basic details related to the customer and exploit the system.
6. **Physical security**: Physical exposure of the network components and smart devices are vulnerable to intrusion. If devices are not properly physically secured, anyone can connect to them to change the network settings or may perform any malicious activity.

There are several ICT-based smart grid components that also suffer from different cyber-security issues.

1. **Operational systems**: Meter Data Management System (MDMS), Supervisory control and data acquisition (SCADA) Systems, Utility system, etc. For example, the system may have several open ports that can be exploited by the attacker.
2. **Standard IT systems**: PCs, servers, mainframes, application server, database server, web server, etc. For example, there is a possibility that apart from port numbers generated by specific applications other port numbers can be open, which can be exploited by the attacker.

3. **Terminating devices**: Smartphones, electric vehicles, smart meters, and other mobile devices. For example, smart meters can be hacked to increase or decrease power demand.

## 5 Possible Solutions of the Cyber-Security Issues

It is evident from the previous section that the smart grid is vulnerable to different types of cyber-security issues. To deal with these vulnerabilities, prevention is considered to be the most effective strategy as compared to elimination. This section provides a possible solution to deal with various cyber-security issues mentioned in different areas of the previous section [5, 9, 13].

1. **Architecture**: It must be designed so that they can be able to handle malicious attacks like denial-of-service (DOS). The network must also be able to cope-up with the network failures oriented attacks by maintaining the automation service locally.
2. **Communication protocols**: To make communication secure, end-to-end encryption must be used. Secure wireless protocols like WPA2 can be used for securing data in wireless networks.
3. **Interfaces**: Web interfaces can be secured by following one or many of the given means: context output encoding, secure password storage, multi-factor authentication, etc. Context output encoding is a programming technique that can prevent cross-site scripting flaws.
4. **Home area networks**: Wireless communications between smart appliances and central systems should be secured by using encryption. Also, there is a need to eliminate rogue access devices to protect against interception or manipulation.
5. **Human factors**: The devices should be capable to use encryption for achieving authentication and authorization to prevent sniffing password attacks. The process of creating user-id and password should be complex to thwart any dictionary-based attacks.
6. **Physical Security**: Providing physical security of the entire network including connected devices is majorly a daunting task. However, infrastructure at the grid, sub-station, and (if possible) at the customer level must be protected by some means of physical guarding and surveillance guarding for physical security.

The smart grid's ICT components are also needed to be resolved for vulnerability so that the cyber-security issues related to them can also be prevented.

1. **Operational Systems**: The operational systems may include one or many of the components like wireless communication, data collection and management, and utility system. The means of securing each one of them is already given in the discussion above.
2. **Classic IT systems**: The IT systems especially web servers, application servers, the database server can be made more secure by resolving the vulnerabilities related to their operating systems, applications, protocols, and network. The

found vulnerabilities can be resolved by updating them through the required patch given by the solution provider.

3. **Terminating devices**: The terminating devices must be purchased from the authorized center, and their operating system and applications must be updated regularly for achieving security.

## 6 Conclusion

Smart grid technology is a recent research area in which issues of existing grid infrastructure are addressed. The smart grid is basically used to monitor various grid-oriented activities, load side preferences, and to perform real-time actions. The smart grid consists of various distributed and heterogeneous computer systems, required to integrate the various forms of energy and to deliver electricity more easily to the consumers. Despite the various advantages of the smart grid, there are multiple challenges in its implementation which include coordination and adoption of the new technology. This article focuses on identifying different cyber-security issues about the current state of a smart grid. These important issues need to be addressed to make the smart grid implementation successful. More specifically, the paper highlights the cyber-security issues in key areas of architecture, communication protocol, interfaces, home area networks, human factors, physical security, and different ICT components. The study finally gives the state of the art solutions to deal with the mentioned cyber-security issues for improving smart grid security.

## References

1. Zhou J, He L, Li C, Cao Y, Liu X, Geng Y (2013) What's the difference between traditional power grid and smart grid?—From dispatching perspective. In: IEEE PES Asia-Pacific power and energy engineering conference (APPEEC)
2. Bari A, Jiang J, Saad W, Arunita J (2014) Challenges in the smart grid applications: an overview. Int J Distrib Sens Netw 10(2)
3. Chebbo M (2007) EU smart grids framework "Electricity networks of the future 2020 and beyond". In: IEEE power engineering society general meeting, Tampa, pp 1–8
4. Tuballa ML, Abundo ML (2016) A review of the development of smart grid technologies. Renew Sustain Energy Rev 59:710–725
5. Mrabet ZE, Kaabouch N, Ghazi NE, Ghazi HE (2018) Cyber-security in smart grid: survey and challenges. Comput Electr Eng 67:469–482
6. Wang W, Lu Z (2013) Cyber security in the smart grid: survey and challenges. Comput Netw 57(5):1344–1371
7. Bigerna S, Bollino CA, Micheli S (2016) Socio-economic acceptability for smart grid development—a comprehensive review. J Clean Prod 131:399–409
8. Kappagantu R, Daniel SA (2018) Challenges and issues of smart grid implementation: a case of Indian scenario. J Electr Syst Inform Technol 5(3):453–467
9. Baumeister T (2010) Literature review on smart grid cyber security. Collaborative Software Development Laboratory at the University of Hawaii

10. Wang W, Xu Y, Khanna M (2011) A survey on the communication architectures in smart grid. Comput Netw 55:3604–3629
11. Yang Y, Littler T, Sezer S, McLaughlin K, Wang HF (2011) Impact of cyber-security issues on smart grid. In: 2nd IEEE PES international conference and exhibition on innovative smart grid technologies, Manchester
12. Pearson ILG (2011) Smart grid cyber security for Europe. Energ Pol 39(9):5211–5218
13. Yan Y, Qian Y, Sharif H, Tipper D (2012) A survey on cyber security for smart grid communications. IEEE Commun Surv Tutorials 14(4):998–1010
14. Aillerie Y, Kayal S, Mennella JP, Samani R, Sauty S, Schmitt L (2013) White paper: smart grid cyber security. Intel, ALSTOM and McAfee
15. Berthier R, Sanders WH, Khurana H (2010) Intrusion detection for advanced metering infrastructures: requirements and architectural directions. In: First IEEE international conference on smart grid communications
16. Sou KC, Sandberg H, Johansson KH (2013) On the exact solution to a smart grid cyber-security analysis problem. IEEE Trans Smart Grid 4(2):856–865
17. Abdulatif A, Kumarage H, Khalil I, Atiquzzaman M, Yi X (2017) Privacy-preserving cloud-based billing with lightweight homomorphic encryption for sensor-enabled smart grid infrastructure. IET Wirel Sens Syst 7:182–190
18. Yu X, Cecati C, Dillon T, Simões MG (2011) The new frontier of smart grids. IEEE Ind Electron Mag 5(3):49–63
19. Setiawan AB, Syamsudin A, Sasongko A (2015) Implementation of secure smart grid as critical information infrastructure in Indonesia: a case study in smart grid electricity. In: Fourth international conference on cyber security, cyber warfare, and digital forensic (CyberSec), pp 34–39

# Data Streaming Architecture for Visualizing Cryptocurrency Temporal Data


**Ajay Bandi**

**Abstract** The utilization of data streaming is becoming essential in mobile computing applications to reduce latency and increase bandwidth. Vast amounts of data are generated continuously from the Web sites of stock markets and financial institutions. The data's meta-analysis is critical for investors and needs to analyze in a short time. Traditionally, it requires several heterogeneous resources with high storage capacity to process and compute the data. Data streaming helps to capture, pipeline, and compute the data without storing it. This research aims to visualize the continuous updates to the cryptocurrency temporal data using aggregations and simple response functions. The cryptocurrency data is collected from multiple data sources. A macro-enabled Excel external live data from Web feature, `C3.js`, and Tableau tools are used to capture and pipeline the streamed data in real time to make better decisions. The results show that the visualizations are dynamically updating in the events of trades in cryptocurrencies over time. Data streaming researchers and practitioners benefit from extending the streaming architecture methodology and dataflow to other domains.

**Keywords** Mobile edge computing · Data streaming · Cryptocurrency · Visualization · Architecture · Big data

## 1 Introduction

In the context of big data, "data streaming" means a continuous generation of data from various data sources. It is also referred to or interchangeably used as streaming data [13]. In various domains, it is essential to observe and react to events in real time. Events such as analyzing the trends for investing in stock markets, identifying unauthorized transactions in bank accounts, detecting cyber-attacks, analyzing social media posts for privacy and intelligence purposes, and monitoring sensor data to detect wear out parts in machinery are a few examples.

A. Bandi (✉)
Northwest Missouri State University, Maryville, MO 64468, USA
e-mail: ajay@nwmissouri.edu



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_50

The streaming solutions are necessary for high-speed and voluminous data. Data streaming is essential for mobile edge computing (MEC), an alternative to centralized cloud computing, to reduce latency [10]. In MEC, the storage and computation happen on the Internet's edge of mobile devices to overcome network communication challenges such as latency and bandwidth. The leading global market intelligence of communications and technology, International Data Corporation (IDC), forecasts that there will be 42 billion IoT devices, and the estimated data generated from these devices is 80 zettabytes ($8x10^{22}$) by the year 2025 [14]. Storing and computing such voluminous real-time data are highly challenging. Data streaming benefits the capturing, process, and bringing insights into massive data without storing it to improve MEC latency. The data stream processing requires latency in milliseconds to seconds using simple response functions and aggregation techniques. In contrast to traditional data warehousing, data streaming uses non-relational data to capture, pipeline, and then to compute the results using simple functions.

The continuous data generated from mobile applications [1] such as IoT apps, sensors, and vehicular systems makes it hard to capture, process, analyze, and get insights from real-time data without storing it. In some instances of data streaming, responding to the events can be done automatically. However, there is a need for human intervention to detect the changes, especially in the temporal data, when there is a continuous generation of vast amounts of data. To overcome this, there is a need to visualize the data in real time to monitor trends. In this paper, the proposed method is to visualize the real-time temporal data of changes in different cryptocurrencies' prices using macro-enabled Excel and visualization tools. The rest of this manuscript is organized as follows: Sect. 2 presents the related work. Section 3 describes the methodology and the implementation of the visualization of the streaming data. Section 4 presents the results, and Sect. 5 concludes with future work.

## 2 Related Work

Data streaming has wide variety of applications in various domains. Ehrlinger et al. [6] discussed industrial streaming data. They conducted a case study in a production plant that generates huge amounts of data. Their study dealt the stability of the production process using machine learning algorithms to handle semantic shift of streaming data. The literature of the data streaming architectures concentrated on reducing latency, accuracy, efficiency, and resource allocation. Henning and Hasselbring [7] focused on reliability and scalability. They proposed a data stream aggregation-based architecture to protect the scalability and reliability of the streaming process. The authors defined these quality attributes' requirements that can be considered for the multilayer and multi-hierarchical systems. The suggested architecture is applied to an industrial case study and observed a linear relationship with the sensor data and reliability.

Tiburtino et al. [12] presented a new method called Xtreaming. This method aims to update the visualizations without visiting multidimensional data exactly once continuously. Yang et al. [15] and Ben-Elizer et al. [5] addressed the robust streaming algorithms in the insertion-only model. Ragan, Stamps, and Goodall [13] proposed a focus and context awareness to visualize the streaming data with visual aggregation technique. However, the data resolution was diminished to represent the context for a longer time. Their results show a negative impact for the contextual awareness in visualization of streaming data. With the emergence of cloud [11] and edge computing [4] applications, there is a need for data processing and visualizing without storing it. The data streaming aggregation and append-only techniques are used to visualize the cryptocurrency temporal data.

## 3 Methodology

The data processing lambda architecture is used to visualize the real-time change in cryptocurrency values [8]. The three main components of a lambda architecture [9] are the batch layer, speed layer, and service layer, as shown in Fig. 1. The new data is fed to both the batch and speeding layers simultaneously from multiple data sources. These sources include but are not limited to IoT applications, mobile apps, vehicle communication with moving parts of traffic, satellites, sensors, and other devices. This speed component includes the inbuilt external streaming data function of the macro-enabled Excel. The batch layer has the complete data that is immutable, append-only and serves as the historical data.

The speed layer contains the most newly added data not yet thoroughly recorded and classified by the serving layer. This layer holds the data that the service layer is currently recording and new data that arrived after the current indexing. It is common to notice the delay in the latest data added to the system and the same data queried by the service layer—technologies such as Apache Spark, Flink, and Amazon Kinesis are used to reduce the latency between the speed and service layers. The service layer contains the regular additions of indexed data and is queriable for the end users to visualize and perform predictive analytics. The standard queries' results are also updated to the complete data in the batch layer and direct usage to end users.

The research goal is to visualize the streaming data. This research proposed the dataflow diagram, as shown in Fig. 2, to visualize the continuous real-time data using lambda architecture. The horizontal cylinders represent the streamed data. The next subsections describe the steps involved in the proposed dataflow diagram.

### *3.1 Tools Used*

To implement a visualizing streaming data project, the following tools are used.

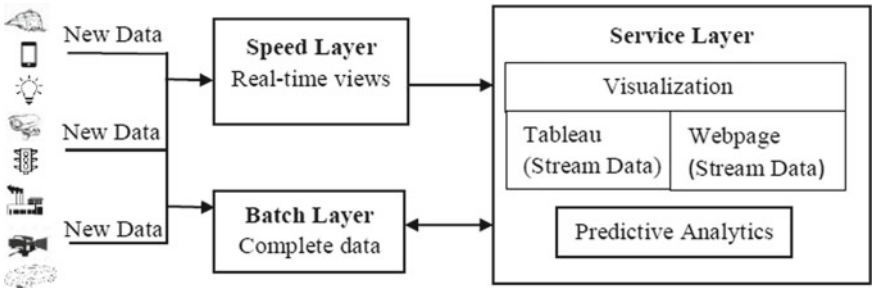



**Fig. 1** Lambda architecture for data streaming

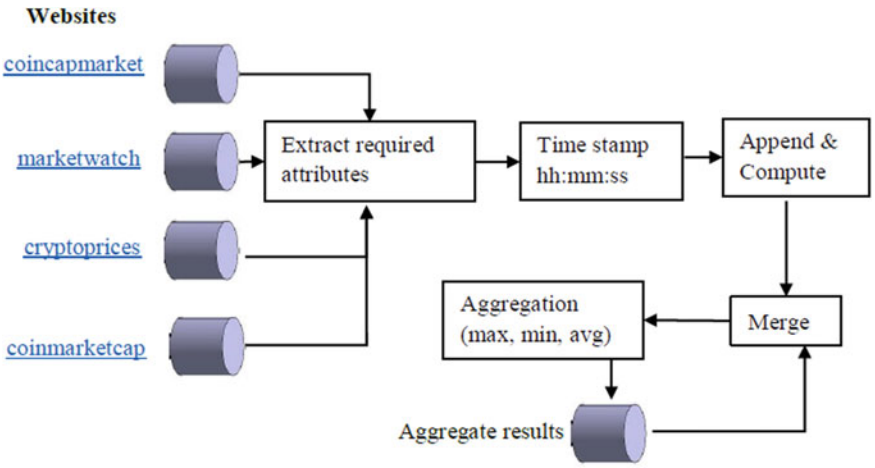



**Fig. 2** Dataflow diagram of data streaming and computation

- Macro-enabled Excel—To extract live data from the Web sites—VBA and live data extraction. Visual basic for applications (VBA) Script is used to curate data, perform aggregation operations, conservation of Excel to `.csv` file, and connect it with web pages for visualizations.
- Papa Parse—It is a fast and powerful CSV parser for the browser that supports Web developers and streaming large files.
- `C3.js`—D3-based reusable chart library for customized visualization purpose on the web pages
- Tableau—A business intelligence and visualization tool to visualize the temporal data using line and multiline charts.

### *3.2 Data*

The different cryptocurrencies such as Bitcoin, Ethereum, Tether, XRP, Chain Link, Bitcoin Cash, Binance Coin, and Cardano, Litecoin, Polkadot, Wrapped Bitcoin, USD coin, and DOW Jones industrial average's stock price are considered in this research. Also, the change in the percentage of the current price of these currencies is used for computations.

### *3.3 Data Sources*

The cryptocurrencies and the stock price are colletecd from multiple data sources. The data is scraped from multiple Web sites (coincapmarket.com, marketwatch.com, coinmarketcap.com, cryptoprices.com). These Web sites are chosen to collect similar attributes related to the goal of this research project. The extracted data is cryptocurrencies and the stock price that changes in real time whenever there is a change in the market price.

### *3.4 Data Extraction*

A macro-enabled spreadsheet (Microsoft Excel) is used to extract the data. The price of Bitcoin, Ethereum, Tether, XRP, Chain Link, Bitcoin, Bitcoin Cash, Binance Coin, Cardano, Litecoin, and Polkadot is extracted from the coincapmarket.com Web site. The Dow Jones industrial average's stock price is from marketwatch.com Web site. The USD coin currency value is from the coinmarketcap.com Web site. The Wrapped Bitcoin's value obtains from the cryptoprices.com Web site. The data from the four different Web sites is extracted into the four different Excel sheets in one Excel workbook. Also, the change in percentage of the current price is extracted and compared to the price before 24 h. This attribute is used to perform the calculations on the streamed data.

To collect the real-time data, embed the Web site URL by clicking the Data tab under the "Get External Data" group, select "From Web." Provide the Web site URL in the "New Web Query" window to import the Web site's raw data. Similarly, import the raw data from other Web sites in separate sheets. Identify the selected cryptocurrency values and stock prices from the raw datasheets. The new sheet extracts the selected values from the four sheets using Excel formulae (Example: Sheet1!D155). The data in the new sheet is the cleaned data and is later used for the visualization. Then, set up the connection properties for live reloading for every one minute.

### 3.5 Insert Time stamp

Since the cryptocurrencies change dynamically in real time according to the market trend, the time stamp is included by using the visual basic for applications (VBA) script within the Excel using the shortcut (Alt+F11). The time is represented in the hh:mm:ss format in a 24 h clock. The time stamp is an essential attribute for the temporal data to represent the changes. The significance of the time stamp in this research is to identify the unique record of the data and the corresponding cryptocurrency price.

### 3.6 Append and Compute

After executing the above steps, the initial values of cryptocurrencies and stock prices are recorded in the respective columns' first row. The initial values are the values extracted from the Web site for the first time after execution. Then, append the values in the new row for every one minute. The new row may have two possible values. (1) The new price of the cryptocurrency or the stock is based on the market. (2) If the value is not changed, the previous value will be recorded at a new time stamp. In this way, the live stream data is recorded and appended in the Excel for further computations and analysis. In other words, the first column is filled with the initial values of all the cryptocurrencies. The second column is the time stamp. From the third column onward, consider the transpose of the first column cryptocurrencies to stream the given time stamp's respective values. For example, if the first column has ten cryptocurrency values, there will be ten respective columns from column #3 to #13. The pseudocode for appending streaming data is given in Algorithm 1.

**Algorithm 1:** Pseudo code for appending streaming data

```
Declare and initialize int LastFilledCell ← address of last filled cell in a
column
    ▷ N is total number of different crypto currencies or the number of values in
    the column one
for int i=0; i<=N.size; i++ do
    if LastFilledCell.value != N[i] then
        Append to LastFilledCell+1.value = N[i].value;
End
```

The change in the value of the cryptocurrency price is calculated based on the change in percentage value. The result is the value of the cryptocurrency before 24 h. The change in percentage value is streamed on the Web site and extracted during the data extraction phase. The calculated change in the value of the price will help the investor for better decision making. For example, the value of the Bitcoin at 14:20:00 is $20,000, and the percentage change is +2% compared to the previous day at the same time. The change in value is $400.

### 3.7 Merge and Aggregation

While capturing the Web sites' streaming data, aggregation functions are used and calculated the average, maximum and minimum values of the Bitcoin and the time stamp. In the VBA script, average, max, min functions of the `WorksheetFunction` class are used to calculate aggregation values. The computed values are merged in the original data and compared with newly added data to calculate the final results for streamed data.

### 3.8 Aggregation Results

The results derived from the aggregation step are compared with the newly streamed data to calculate the new maximum and minimum values. Consider the overall past data to recalculate the new average of the streamed data entry. The VBA pseudocode for aggregation results is given Algorithm 2.

**Algorithm 2:** Pseudo code for calculating aggregations on streaming data

```
Create a column named max in Excel
Declare and initialize int MaxColLastFilledCell ← address of last cell in max column
                                                  ▷ N is total number of different crypto currencies
                                                  ▷ column is size of a respective column
for (int i=0; i<=N.size; i++) do
    for (int j=0; j<=column.size; j++) do
        if MaxColLastFilledCell.value < N[i] then
            Append MaxColLastFilledCell.value = Max(column)
            Get timestamp and append

                                                  ▷ to find min
int MinValue ← 9999.
                                                  ▷ this will be declared in excel in min column cell 1
Declare and initialize int MinLastFilledCell ← address of last cell in min column
                                                  ▷ N is total number of different crypto currencies
                                                  ▷ column is size of a respective column
]for (int i=0; i<=N.size; i++) do
    for (int j=0; j<=column.size; j++) do
        if MinLastFilledCell.value > N[i] then
            Append MinLastFilledCell.value = min(column);
            Get timestamp and append

        ▷ To find Average Declare and initialize int AvgLastFilledCell ← address of last cell in average column
Append AvgLastFilledCell.value = Average(column)
                                                  ▷ column is size of a respective column
```

## 4 Results and Discussion

The cleaned data set is used in the speed layer for real-time views and then pipelined to the service layer. Visualizations and prediction analytics are performed in the service layer. For the real-time visualization of the cryptocurrencies' temporal data, the first

step is converting the macro-enabled Excel sheet into `.csv` files. The pseudocode for the conversion in VBA script is given in Algorithm 3. Then, the cleaned data in the .csv file is used to visualize on the web page. Papa Parse is used to parse the `.csv` files and `C3.js` to visualize. Papa Parse library is a fast and powerful CSV parser for the browser that supports Web workers and streaming large files. It is easy to use and parse `CSV` files directly to local or over the network. C3 is a JavaScript library that builds on top of D3. C3 makes it easy to generate D3-based charts by wrapping the code required to construct the entire chart. The pseudocode for visualizing streaming data using Papa Parse and `C3.js` is given in Algorithm 4.

**Algorithm 3:** This is a pseudo code for converting .xlsm to .csv file

```
Start:
Declare Excel.Worksheet ws;
Declare String SaveToDirectory;
Initialize SaveToDirectory as "Streaming-Visualization \ Web Visualization \data"
foreach ws in ThisWorkbook.Worksheets do
    ws.SaveAs SaveToDirectory & ws.Name, xlCSV
end
End
```

**Algorithm 4:** This is a pseudocode for visualizing streaming data.

```
Start:
functionparseData(createGraph)
Read the CleanedDataset.csv using Papa Parse library
Call createGraph (CleanedDataset.data)
End Function
functioncreateGraph(data)
                                        ▷ data is excel data from CleanedDataset.csv
Initialize time of an array type.
Initialize array variables for all the cryptocurrencies in CleanedDataset.csv
for var i = 1; i < data.length-1; i++ do
    Add the time stamp from the CleanedDataset.csv to time array
    Add cryptocurrencies data into its respective arrays
end
Generate multiline chart using c3 library taking time on X-axis and cryptocurrencies on Y-axis
functiontimedRefresh(timeoutPeriod)
                ▷ timeoutPeriod is the period of time in milliseconds for refreshing the webpage
                                ▷ Web page will reload for every 60 seconds i.e. 6000ms
End Function
End
```

The visualizations of the streaming data are shown in Figs. 3 and 4. These charts are dynamic, and the visualizations update every one minute. The web page is developed to automatically refresh every minute to update the data dynamically in the visualizations. The change in the cryptocurrency prices is temporal data as it changes according to the market trend. The resulted visualizations compare various cryptocurrencies in a multiline chart, and the Bitcoin's price change is presented in the line chart. The user can filter whatever the cryptocurrency they want to view. Figure 4 shows the line chart of Bitcoin's trend after filtering it from other cryptocurrencies. The real-time visualization can be viewed in the video [2], and the project artifacts can found in the GitHub repository [3].

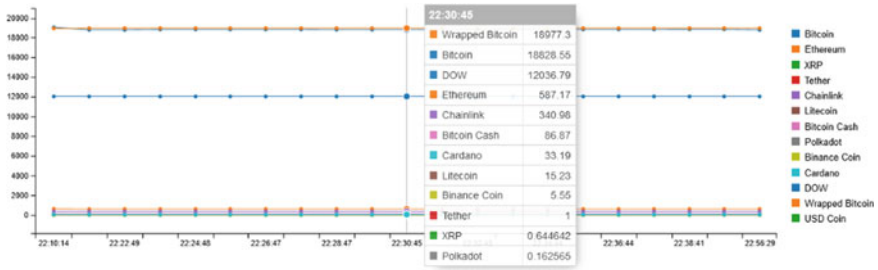



**Fig. 3** Multiline chart for streamed data on a web page


**Fig. 4** Line chart for Bitcoin streamed data on a web page

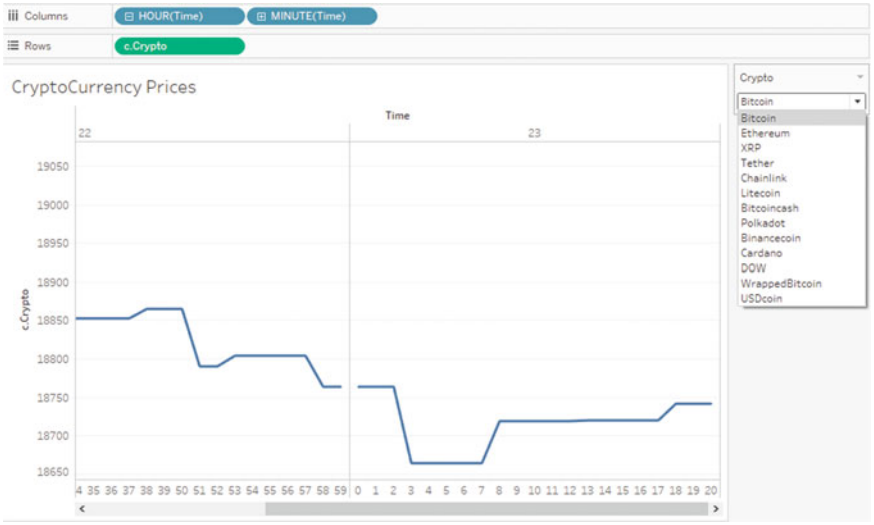



**Fig. 5** Line chart for Bitcoin streamed data in tableau

On the C3 chats, the Y-axis represents the dynamic increment of the time stamp, and it changes as the new data arrived for every minute. The *X*-axis represents the cryptocurrency value that automatically adjusts based on the change in the prices. The tool tip in Figs. 3 and 4 shows the time stamp and the corresponding cryptocurrency

values at that instance. In tableau visualizations, the $X$-axis represents the time for every minute (0–59) for every hour. The $Y$-axis represents the cryptocurrency value. In Fig. 5 the line chart shows from 22:35 to 23:20. After 22:59, the new hour 23 started and continued to represent the data for every minute to visualize the streaming data's dynamic changes.

## 5 Conclusions and Future Work

Software systems that analyze the continuously generated data, also called streaming data, need to visualize in real time to make better decisions. This research presented a method to visualize cryptocurrencies in that lambda architecture for data streaming. The implementation method uses aggregations, append-only and simple response functions of data streaming for temporal data. The results show that the visualizations are updating for every minute dynamically. The implementation of the architecture and the artifacts are provided in the GitHub repository. In the future, this research will be extended to measure and reduce the latency along with the security aspects of the data streaming.

## References

1. Bandi A, Fellah A (2017) Design issues for converting websites to mobile sites and apps: a case study. In: 2017 international conference on computing methodologies and communication (ICCMC), pp 652–656. https://doi.org/10.1109/ICCMC.2017.8282547
2. Bandi A. https://youtu.be/Ivc6X4KalRc
3. Bandi A. https://github.com/bandiajay/DataStreamingVisualization
4. Bandi A, Hurtado JA (2020) Edge computing as an architectural solution: an umbrella review. In: 26th annual international conference on advanced computing and communications. Springer
5. Ben-Eliezer O, Jayaram R, Woodruff DP, Yogev E (2020) A framework for adversarially robust streaming algorithms. In: Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI symposium on principles of database systems, pp 63–80
6. Ehrlinger L, Lettner C, Himmelbauer J (2020) Tackling semantic shift in industrial streaming data over time, pp 36–39
7. Henning S, Hasselbring W (2020) Scalable and reliable multi-dimensional sensor data aggregation in data streaming architectures. Data-Enabled Discov Appl 4(1):1–12
8. Horvat N, Ivković V, Todorović N, Ivančević V, Gajić D, Luković I (2020) Big data architecture for cryptocurrency real-time data processing
9. Lin J (2017) The lambda and the kappa. IEEE Internet Comput 21(5):60–66. https://doi.org/10.1109/MIC.2017.3481351
10. Mahadev S (2017) The emergence of edge computing. Computer 50(1):30–39
11. Maraparedy R, Bandi A, Tirumala SS (2012) Cloud computing architectures: a retrospective study. J Innov Comput Sci Eng 2(1):1–5
12. Neves TT, Martins RM, Coimbra DB, Kucher K, Kerren A, Paulovich FV (2020) Xtreaming: an incremental multidimensional projection technique and its application to streaming data. arXiv preprint arXiv:2003.09017

13. Ragan ED, Stamps AS, Goodall JR (2020) Empirical study of focus-plus-context and aggregation techniques for the visualization of streaming data. In: Proceedings of the international conference on advanced visual interfaces, pp 1–5
14. Shirer M, MacGillivray C The growth in connected IoT devices is expected to generate 79.4ZB of data in 2025, according to a new idc forecast. https://www.idc.com/getdoc.jsp?containerId=prUS45213219
15. Yang R, Xu D, Cheng Y, Wang Y, Zhang D (2020) Streaming algorithms for robust submodular maximization. Discrete Appl Math

# An Overview of Layer 4 and Layer 7 Load Balancing


**S. Rajagopalan**

**Abstract** Load Balancing is a significant process to dispense the numerous types of traffic into various servers and clients. The key aim of load balancing is every single server must not be overloaded with tasks. Disseminate the network traffic through various routes would reduce the congestion in the network and it would reduce the latency of the servers. Actually, load balancing ensures the server's well-being. There is a high possibility of a server fall through when there is no load balancers assigned to servers. Load balancers enhance the overall proficiency of the servers and minimizing the load of the server by efficient traffic management. The OSI reference model characterizes four types of load balancing namely application, transport, network, and channel. This review article intends to analyze the pros and cons of layer 4 and layer 7 load balancing in the OSI reference model.

**Keywords** TCP (Transmission control Protocol) · UDP (User datagram Protocol) · HTTP (HyperText transfer Protocol) · HTTPS (Hypertext transfer protocol Secure) · OSI (Open systems Interconnection) · SMTP (Simple mail transfer Protocol) · ADC (Application delivery Controller) · Offloading

## 1 Introduction

In the present network scenario, websites are forced to handle numerous requests from clients. When a data center nears or reaches or exceeds the threshold level then it is considered a congested network [1]. A pool of servers must send back the proper images, audios and videos for these client requests. Mobile users receive a variety of services from clouds [2]. In order to satisfy the customer needs, more servers must be added. Load balancers behave like a traffic regulator just before the server to route all the client's requests [3]. The application delivery controller distributes the network traffic optimally across over used servers to less-used servers [4]. The vital function of ADC is load balancing. Load balancing can be done in different

S. Rajagopalan (✉)
Department of Computational Logistics, Alagappa University, Karaikudi, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_51

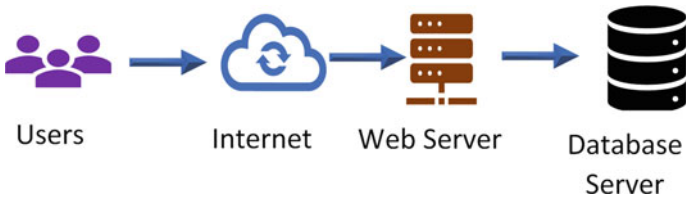



**Fig. 1** User—web server connection without load balancing. *Source* https://www.esds.co.in/blog/types-of-load-balancing/

ways at layer 4 as well as layer 7 [5]. This layer 4 and layer 7 load balancing is based on the OSI reference model. Layer 4 uses TCP and UDP protocols and uses simple load balancing algorithms to balance the traffic among networks in order to avoid congestion [6]. It uses basic information like server connections and response time. Layer 7 load balancer works at the topmost layer of OSI reference model, i.e., it works on the application layer. Enhanced information known as header, content, cookies, URLs, and HTTP allows it to make the correct routing decision. Network services are utilized to manage network traffic and also provide data storage, manipulation, and communication services. Cloud provides a lot of storage area in a secured way [7]. Let us see the differences between Layer 4 and Layer 7 in forthcoming sessions.

## *1.1 Need for Load Balancing*

Figure 1 shows the client-server connection without load balancing. Scalability is the principal problem in load balancing, despite many load balancers are introduced at several levels. Sometimes, sites may not be accessible due to imbalance load distribution. So obtaining optimal load and distribute it by using routing algorithms to achieve optimal functioning is still a challenge. Cloud bursting is one of the techniques to resolve load balancing issues in the cloud environment [8].

## *1.2 Variances of L7 and L4 Load Balancing*

Layer 4 load balancing optimizes the traffic at the transport lever. It uses information such as ports and protocols to manage the traffic without checking the actual content of the messages. For simple packet-level load balancing this layer 4 load balancing is enough. Poor utilization of the resources may lead to congestion, so load balancing must be done properly [9]. It uses simple algorithms such as round-robin. It doesn't have the visibility of the content. Layer 4 never inspect the content or decrypt them to forward quickly as well as in a secure manner. So routing will not be based on media type, localization rules, etc.

It operates at the application layer. L7 load balancer uses protocols such as HTTP, HTTPS and SMTP. L7 finalizes the routing with respect to the message's subject. Layer 7 load balancer inspects messages and the routing decision is taken based on the content of the message. It can terminate traffic, perform decryption and provide end-to-end security. L7 load balancer creates new TCP connection to the upstream server. TCP may have issues in high-speed networks due to packet loss [10].

The need for encryption reduces the performance at Layer 7. But this can be almost neutralized by the user of SSL offload functionality. L7 load balancing enables context sensible networking. It increases server efficiency by sending all client request to the same server. The visibility at the packet level allows content caching. L7 load balancer provides intelligence to handle piggyback or multiplex requests onto a single connection. This reduces the overhead and optimizes the traffic.

On the whole, layer 4 load balancing is meant for applications that do not access HTTP and layer 7 load balancing is a request/reply service [11].

### *1.3 ADC Load Balancing*

Layer 7 load balancers offer intelligent routing decisions with more functionality. Its visibility and application awareness enable intelligent routing, optimizations, and performance enhancement. For example, The user can access the correct content version according to the language indicated in the browser header. This gives best experience for any user, device, and site. ADC load balancing from L4 to L7 addresses a variety of client needs for diverse applications.

ADCs provide advanced layer 4 to layer 7 load balancing and provide high availability and business continuity. It enables fast response time with optimal and intelligent traffic distribution by customizing traffic such as blue/green traffic. The SSL/TLS offloading techniques further optimize application performance.

At layer 4 the load balancer delivers the traffic by knowing the limited network information like ports and protocols and uses simple algorithms like round robin. It delivers the traffic by calculating the best destination path.

In layer 7, load balancer having awareness about applications. It can make complex and informed load balancing decisions with this additional information. By knowing the protocol such as HTTP, it can easily identify client sessions based on cookies.

Layer 7 load balancer can balance the traffic based on the content. For e.g. If any request from the client and requesting for images, that request can be diverted to the server having images. This will reduce the load on application server. The load balancing in OSI and TCP reference models are depicted in Fig. 2.

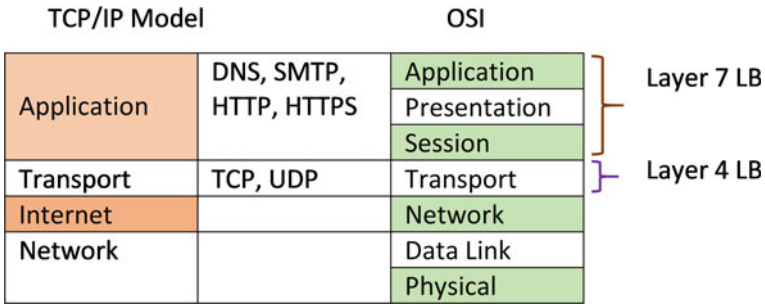



**Fig. 2** Load balancing in OSI and TCP models. *Source* https://freeloadbalancer.com/load-balancing-layer-4-and-layer-7/

## 2 L4 Load Balancing Architectures

In layer 4 load balancer uses TCP and UDP protocols at transport level. Layer 4 load balancer decides the route based on IPs and TCP or UDP protocols. It can view the traffic between client and server as packet. It will take the routing decision for each packet.

In layer 4 connections are established between the client and the server using simple load balancing algorithms. It is fast but it can't do anything on the protocol above layer 4. The fastest layer 4 load balancer uses an ASIC for routing decisions.

List of possible architectures for L4 load balancing are Network Address Translation (NAT), IP Tunnel and Direct Server Return.

### *2.1 Network Address Translation (NAT)*

Client and server packet transmission routing will be done by the load balancer in Network Address Translation (NAT) mode. In this client-server network, a TCP connection is created to forward the network load to the intended server. A server is selected after a group of servers, then it will send the packet to the newly identified server with an altered address. The TCP connection and Dataflow are shown in Fig. 3.

**Employment of NAT** In a transmission NAT can be implemented, when a response time needs to be intensified when there is no necessity for response time, where there won't be any bottleneck in the future, and where there is a need for a default gateway of the servers to be changed.

**Strengths** Load balancing will be done in a faster manner. The deployment is very straightforward and simple.

**Weaknesses** For reverse NAT procedures, server entry points need to use load balancers. According to the load balancer output capacity, the output bandwidth is limited [12].

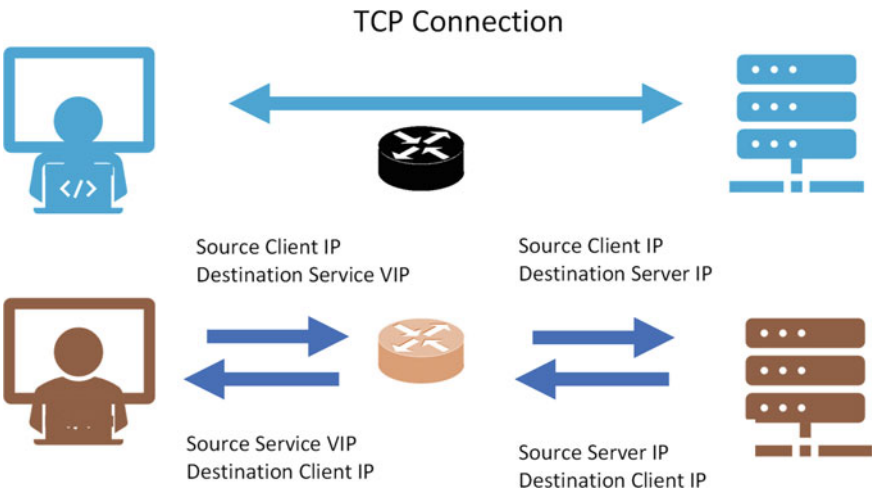



**Fig. 3** NAT: TCP connection and data flow. *Source* https://www.haproxy.com/blog/layer-4-load-balancing-nat-mode/

## *2.2 Direct Server Return (DSR)*

In Direct Server Return (DSR), the MAC address will be changed by the load balancer. It won't change anything except MAC. After this point, client and server will have direct contacts and there is no need for any load balancer. The TCP connection and Dataflow are shown in Fig. 4.

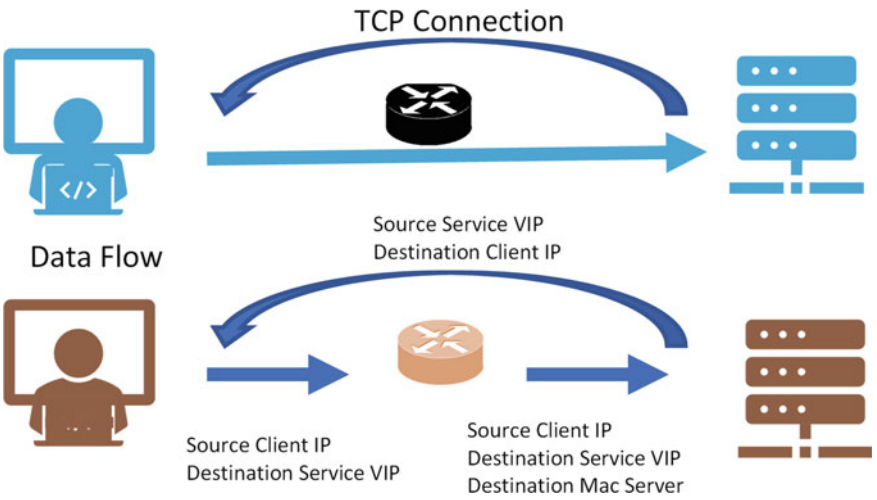



**Fig. 4** DSR: TCP connection and data flow. *Source* https://www.haproxy.com/blog/layer-4-load-balancing-direct-server-return-mode/

In layer 4 load balancing, the TCP connection is established at the backend. But the request passes through the load balancers. The requests by clients are seen by the load balancers and the destination address will be changed.

**Employment of DSR** DSR implementation is also similar to the NAT.

**Strengths** Load balancing is performed in the fastest mode. The Sum of each backend bandwidth is the total output bandwidth. Bottleneck will not be created due to the bandwidth of the load balancer. Less intrusive as compared to the layer 4 load balancing.

**Weaknesses** Service VIP must be configured on each system looping boundary. It should not answer ARP requests. There are no layer 7 advanced characteristics [13].

## *2.3 IP Tunnel Mode*

IP Tunnel mode and DSR mode are somewhat similar up to a certain extent. The users' requests are processed by the master when it receives from the load balancer. Afterward, the clients will receive a reply straight away from the servers. In IP tunnelling, the load balancer compresses a demand and sends it to the master. IP tunnelling helps to direct the traffic between the service user and the server. The TCP connection and Dataflow are shown in Fig. 5.

**Employment of IP Tunnel mode** Routing is the only way to access the backend servers. The rest of the scenarios is similar to NAT and DSR.

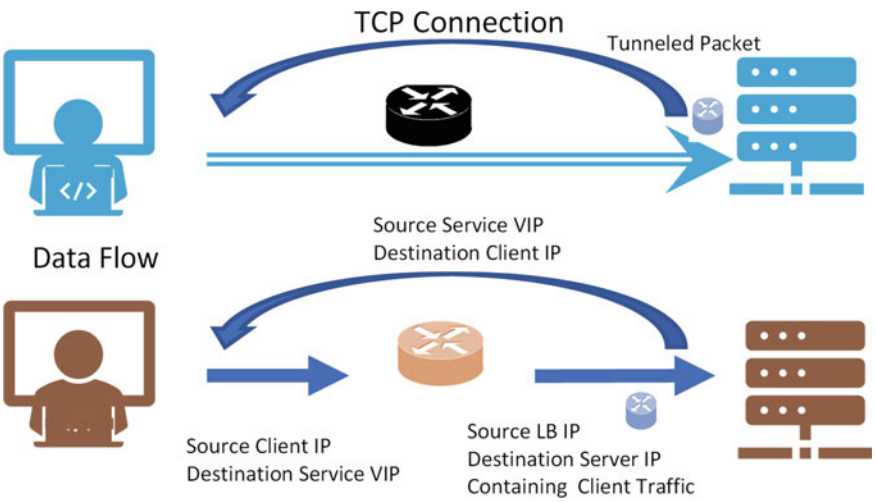



**Fig. 5** IP tunnel mode: TCP connection and data flow. *Source* https://www.haproxy.com/blog/layer-4-load-balancing-tunnel-mode/

**Strengths** Multiple datacenters' backends could be utilized. The other strengths are alike DSR.

**Weaknesses** It requires a patched backend so that it will be able to tunnel the IP traffic. There are no layer 7 advanced features available [14].

## 3 L7 Load Balancing Architectures

Layer 7 is related to the 7th layer of the OSI model, i.e, application level. Example HTTP, FTP, SMTP, DNS protocols. Layer 7 load balancer makes the routing decisions based on IPs, TCP, or UDP ports. It can get any other information from the application protocol (mainly HTTP) and takes the routing decision. Even this kind of process seems to be slow, but actually, it took less than a millisecond.

Proxy mode is the only architecture with two main flavours, namely Proxy Mode and Transparent Proxy Mode.

### *3.1 Proxy Mode*

This is also called as reverse-proxy. All transactions between the user and the server are passing through the load balancer. If a TCP connection is meant for user, here load balancer behaves like a master. So request-reply will be done properly by the load balancers to the clients. If a TCP connection is meant for the server, load balancer behaves like a service user from the master server. Without any modifications, the client-server technology works well. The TCP connection and Dataflow are shown in Fig. 6.

Load balancer upholds two TCP connections as shown in the diagram.

**Employment of Proxy mode** This method can be realized when the app does not bother about client's IP address, when the application layer needs to work smartly, and to secure an application.

**Strengths** There is no direct server access mode, so it will be secured. This method allows protocol authentication and verification. Load balance is possible even in the case of client and server are in the same subnet.

**Weaknesses** Comparatively slow load balancing done in this method than layer 4. Only IP address can be seen by the backend servers [15].

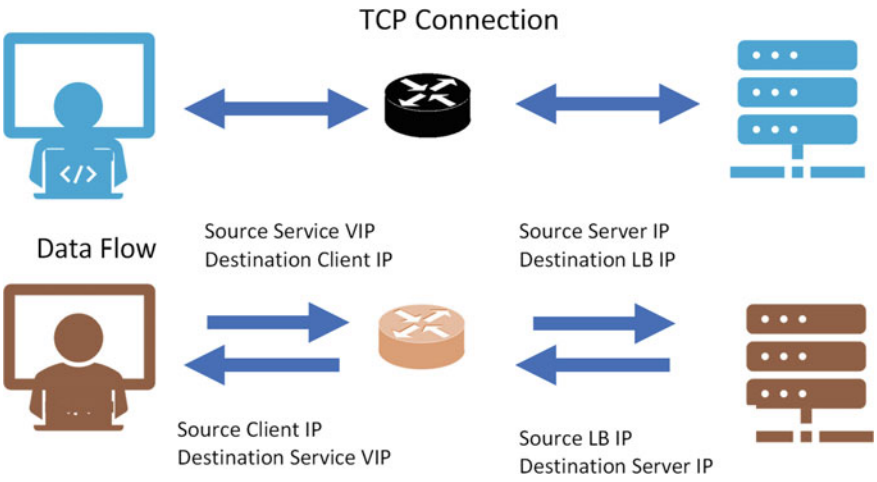



**Fig. 6** Proxy mode: TCP connection and data flow. *Source* https://www.haproxy.com/blog/layer-7-load-balancing-proxy-mode/

### *3.2 Transparent Proxy Mode*

It is almost like the proxy mode, but it differs in a way that the load balancer uses request machine's IP as request origin machine's IP.

A load balancer and two TCP connections are shown in the above figure. Here, server happens to use the load balancer as a default gateway because load balancer commences TCP connection with server by IP address of a client. Or else the server would reply to a client straightaway but client will decline it. The TCP connection and Dataflow are shown in Fig. 7.

**Employment of Transparent Proxy mode** This method can be utilized whenever a network layer level load balancing facility is required for a client IP, when content switching desirable, and to safeguard an application.

**Strengths** Client IP will be notified by server in network layer. The rest are similar to the Proxy Mode.

**Weaknesses** It is a very intrusive method because there must be a variation in gateway of the default server and slower than the layer 4 load balancing. The clients and servers should not be in the same subnet and it must be indifferent subnets [16].

## 4 Conclusion

Layer 4 and Layer 7 load balancing methods, advantages and disadvantages are discussed in this article. Load balancing will improve server uptime. Layer 4 and

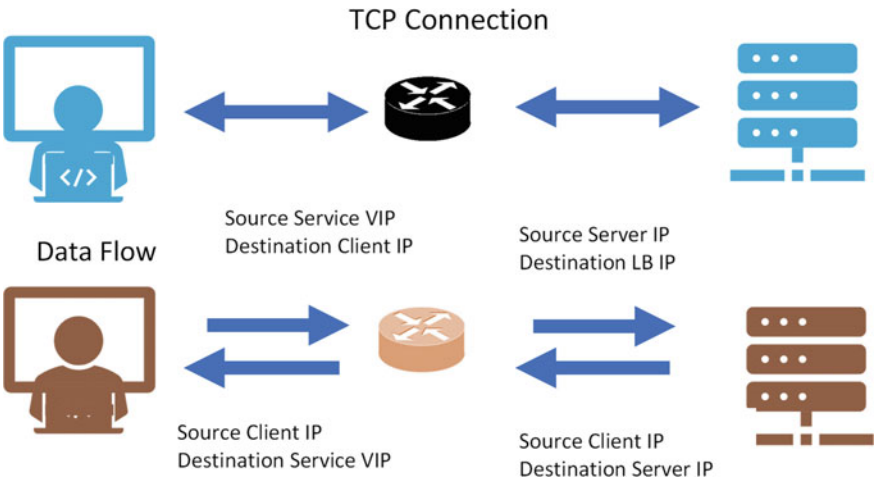



**Fig. 7** Transparent Proxy mode: TCP connection and data flow. *Source* https://www.haproxy.com/blog/layer-7-load-balancing-transparent-proxy-mode/

Layer 7 load balancing are provided by ADC to ensure fast, reliable, and secured communication without any discontinuity. Even a one connection failure may lead to multiple logical connections failure [17]. The primary goal of this load balancing method is to achieve ideal traffic dispersal or personalized traffic dispersal. SSL/TSL computation offloading method impacts positively to obtain the optimal distribution of the network traffic. It's very difficult to ascertain which balancing method is the best to be practiced. It can be decided according to the unique demand of the network traffic and it will resolve the security issues.

## References

1. Rajagopalan S, Naganathan ER, Herbert Raj P Ant colony optimization based congestion control algorithm for MPLS network. In: International conference on high performance architecture and grid computing, HPAGC 2011, vol 169. Springer Berlin Heidelberg, pp 214–223. ISBN No. Online ISSN 978-3-642-22577-2
2. Herbert Raj P, Ravi Kumar P, Jelciana P (2019) Load balancing in mobile cloud computing using bin packing's first fit decreasing method. In: Omar S et al (eds) CIIS 2018, AISC, vol 888. Springer Nature Switzerland AG 2019, pp 97–106. ISBN 978-3-030-03302-6_9, 2019
3. NGINX: what is load balancing? November 2020. https://www.nginx.com/resources/glossary/load-balancing/
4. Herbert Raj P, Ravi Kumar P, Jelciana P, Rajagopalan S (2020) Modified first fit decreasing method for load balancing in mobile clouds. In: 4th international conference on intelligent computing and control systems (ICICCS), 13–15 May 2020, Vaigai College Engineering (VCE), Madurai, India. IEEE
5. Nicholson P (2020) How do layer 4 load balancing and layer 7 load balancing differ? A10 networks, June 2020. https://www.a10networks.com/blog/how-do-layer-4-and-layer-7-load-balancing-differ/

6. Herbert Raj P, Raja Gopalan S, Padmapriya A, Charles S (2010) Achieving balanced traffic distribution in MPLS networks. In: 2010 3rd IEEE international conference on computer science and information technology, July 7 2010, vol 8, Chengdu. Institute of Electrical and Electronics Engineers, Inc. Printed in Beijing, China, pp 351–355. ISBN 978-1-42445539-3
7. Herbert Raj P, Ravi Kumar P, Jelciana P (2016) Mobile cloud computing: a survey on challenges and issues. Int J Comput Sci Inform Sec (IJCSIS) 14(12). ISSN 1947-5500
8. Nugara A (2019) Load balancing in microsoft azure. O'Reilly, Publisher(s): O'Reilly Media, Inc. ISBN 9781492053927
9. Kasmir Raja SV, Herbert Raj P (2007) Identifying congestion hotspots using bayesian networks. Asian J Inform Technol 6(8):854–858. ISSN 1682-3915. Medwell Journals
10. Naganathan ER, Rajagopalan S, Herbert Raj P (2011) Traffic flow analysis model based routing protocol for multi-protocol label switching network. J Comput Sci 7(11):1674–1678. ISSN No. 1549-3636
11. Technical glossary: TCP load balancing. AVI Networks, November 2020. https://avinetworks.com/glossary/tcp-load-balancing/
12. Assmann B (2011) Layer 4 load balancing NAT mode, load balancing/routing. Tech 22 July 2011. https://www.haproxy.com/blog/layer-4-load-balancing-nat-mode/
13. Assmann B (2011) Layer 4 load balancing direct server return mode. Tech 29 July 2011. https://www.haproxy.com/blog/layer-4-load-balancing-direct-server-return-mode/
14. Assmann B (2011) Layer 4 load balancing tunnel mode. Tech 29 July 2011. https://www.haproxy.com/blog/layer-4-load-balancing-tunnel-mode/
15. Assmann B (2011) Layer 7 load balancing proxy mode, load balancing/routing. Tech 3 August 3 2011. https://www.haproxy.com/blog/layer-7-load-balancing-proxy-mode/
16. Assmann B (2011) Layer 7 load balancing transparent proxy mode, load balancing/routing. Tech 3 August 2011. https://www.haproxy.com/blog/layer-7-load-balancing-transparent-proxy-mode/
17. Raja SVK, Herbert Raj P (2001) Balanced traffic distribution for MPLS using bin packing method. In: international conference on intelligent sensors, sensor networks and information processing, 3–6 December 2007, University of Melbourne, Melbourne, Australia. Proceedings are published in the IEEE, pp 101–106, ISBN 978-1-4244-1501-4@ 2007 IEEE. https://doi.org/10.1109/issnip.2007.4496827. Publication Date: 3–6 December 2007, Current Version Published: 2008-04-25

# Integration of IoT and SDN to Mitigate DDoS with RYU Controller


**Mimi Cherian and Satishkumar Verma**

**Abstract** Internet of Things is an upcoming technology, where IoT devices are interacting with cloud over Internet. Large number of IoT devices generate exponential amount of data that creates a huge impact on storage, network elements and specifically on the security and analysis of data. Recent research indicates that there should be a change in the networking paradigm to inculcate the dynamic demands of IoT environment. The network security issue like distributed denial of service [DDoS] attacks are of major concern, and its mitigation at the earliest remains vital. In IoT-related environment, the security issues of traditional network have major impact in IoT application domain. The IoT-related data that are depending on domain of application can be time sensitive or highly confidential, and hence, it arises the need to change the paradigm of traditional network. The expectant network should be more secure and flexible to detect and mitigate the network attacks. IoT environment with software-defined network seems to be promising enough to reduce many security issues with respect to IoT in traditional network environment. The proposed research work has created a test bed that collects IoT live data and sends it through secure SDN into the cloud platform.

**Keywords** SDN · IoT · Security · DDoS · Network

## 1 Introduction

Internet of Things as already known it is an evolving technology, which is having an increasing impact in domains like manufacturing and automation industries, health care, transportation and many more fields. The true potential of IoT can be achieved when existing traditional network paradigm evolves to meet the dynamic requirements of IoT environment. In traditional network, many security issues were

M. Cherian (✉) · S. Verma
Pillai College of Engineering, New Panvel, Mumbai University, Mumbai, India
e-mail: mcherian@mes.ac.in

S. Verma
e-mail: vsat2k@mes.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_52

found like sidechannel attack, denial of service, distribute denial of service, Sybil, blackhole, code injection and many more. The solution for each of these attacks is different and needs to be mitigated separately as traditional network is not flexible enough to adapt to these changes. Hence, improving the existing countermeasures would not be sufficient to resolve security issues in IoT. Recent research suggests to change the approach of resolving IoT network-related issues by implementing proactive measures to resolve network security issues. Many of the security issues in IoT can be reduced if traditional network was more flexible, dynamic and proactive in nature with respect to security and scalability of network [1]. Software-defined network allows IoT network to be managed dynamically also provides flexibility to enhance security in IoT networks by creating applications to prevent, detect and react to threats. SDN decouples the data planes and control plane in a network. Decision-making in SDN is done by control plane, and data forwarding is handled by switches. SDN allows centralized programming of networking and can be managed more easily.

### *1.1 Structuring of Paper*

Section 2 will be the briefing of work done by different researchers that promise the SDN, and IoT integration is new direction of research to change existing network framework for meeting the upcoming demands of IoT applications. The related work has motivated to create an experimental architecture for SDN and IoT integration.

Section 3 will be briefing of work done in DDoS detection and mitigation in SDN environment thus leading to development of a secure SDN framework for IoT.

Section 4 will have architecture of SDN and IoT integration. The subsequent sub-sections are the technical aspects for implementation of IoT test bed, implementation of SDN test bed and setting up cloud platform ThingsBoard is briefed.

Section 5 is simulation of DDoS attack and its mitigation using sFlow-RT to detect DDoS attack, and this flow is given to RYU Controller's Northbound API for mitigation.

## 2 Related Work in SDN and IoT Integration

The related work section will be considering the research papers that are assuring that IoT and SDN integration is promising paradigm for future of IoT. Many researchers have suggested that integration of software-defined network with Internet of Things has great potential [1–3]. The advantages of SDN and IoT integration are recognized in many domains like smart grid settings, smart homes or smart transportation. SDN-IoT integration is also provides security in IoT, because security mechanisms can be implemented easily and network can be more scalable by implementing SDN-IoT integration [4–8].

Jararweh et al. [5] proposed a SDIoT software-defined-based framework model to address the challenges in the traditional IoT architecture to secure and store the data from IoT objects by sending through software-defined network.

The authors in [6] proposed an architecture that applies SDN technology into the IoT such that it improves the security of the IoT. The simulation experiment was provided with the architecture model that can effectively improve the security of the IoT.

The authors in [5] proposed a framework to provide flexible and scalable IoT network that are benefits of SDN. The framework focused on the machine to machine (M2M) transactions.

The work in [7] shows that software-defined techniques can have better network control, especially in IoT networks, such that it can provide better robust and vigorous solutions.

The work in [8] shows the performance difference between SDN-enabled network and traditional networking architecture. This paper helps in designing and implementing SDN test bed for the current research. The literature survey in related work provided research and scope in this domain that has influenced to create a potential experimental test bed for the current research.

## 3 Related Work of DDoS Attack Detection and Mitigation

Distributed denial of service attack can occur in different forms but they all try to flood the network. DDoS can even saturate the network and disrupt from further functioning of network. The attack takes place by flooding of ICMP, UDP and TCP SYN packets. The compromised host attacker will try sending packets to victim host through switch. The controller update the flow table based on request. Once the request is granted, the network is immediately brought down as the attacked host motive was to flood the network toward victim. There are many techniques to mitigate DDOS attack researched and few will be discussed in this section.

In paper [12], authors have discussed emulation of SDN network and in real-time detection and mitigation of DDoS attack using sFlow and Mininet.

In paper [13], authors have used algorithm for measuring entropy that is randomness in network using Shannon's entropy equation. The entropy values change based on DDoS attack. The setup is done using POX controller, Raspberry pi and OpenVSwitch.

The authors of paper [14] have proposed defence mechanisms used are ingress, egress and pushback to validate the legitimacy of packets in DDOS attack. Attacker is identified by applying ingress filtering in same domain and pushback for another domain.

The authors of paper [16] use the concept of entropy and detect whether the network flow is abnormal. BiLSTM-RNN neural network algorithm is used to train dataset thus identifying real-time DDoS attack and reduces overhead on controller.

The authors of paper [17] proposed a secure SDN-based IoT framework. This framework is called SoftThings, and it is used to detect abnormal activities and attacks at the earliest and mitigate it. SDN controller uses concepts of machine learning to monitor the network behaviour. The testing environment is created on mininet emulator.

The authors of paper [6] have utilized the combination of the neural network and the support vector machine. These concepts are used to detect and the classification method for the DDOS attacks in the telecommunication network. NS-2 is used for implementation of the project.

## 4 IoT-SDN Test bed Architecture and Setup

Many research papers in related work section suggest that SDN and IoT integration seems to be promising environment due to flexible and programmable network providing a centralized view of network. The implementation of a experimental test bed for these two unique domains is done in three sections. First section will be creating IoT test bed, second section will be creating SDN test bed, and third section will be integrating both the domains with cloud platform.

The architecture of IoT-SDN test bed is shown in Fig. 1 that has end-to-end connectivity. Starts from collecting the real-time sensor data from live IoT environment and sending the data through SDN network to IoT dashboard on cloud platform Thingsboard.

The Iot test bed will be executed with necessary hardware components. The SDN environment can be executed based on the topology required that consists of RYU controller and respective IoT gateways and hosts.

### *4.1 IoT-Test bed Setup*

The motive of IoT test bed is to capture live sensor data and send it through IoT gateway. IoT test bed can be created using Arduino or Raspberry pi based on scope of the project. The micro-controller is then connected with different sensors as per the project requirement, and currently, it is required to collect the temperature and humidity data; hence, in Fig. 1, the DHT11 sensor is used. The coding for circuit is executed in Fig. 2. These data collected from DHT11 can be used to create .csv file with required amount of live data for dataset.

## *4.2 SDN-Test bed Setup*

The SDN setup is done in an Ubuntu 18 operating system that is installed in a virtual machine. Flat tree topology is created in SDN network which currently consists of IoT gateway, host and controller. The switches used will be OpenFlow-enabled switches that will be connected with RYU controller. The topology is created in mininet. Mininet has better features for simulation of topology as discussed by researchers in related work section. Mininet is integrated with sFlow as sFlow can provide network data statistics. Mininet enables us to create multiple hosts and Web servers that can interact through RYU controller. IoT gateways will have IoT protocols like CoAP or MQTT so that it can transfer data from IoT test bed to IoT cloud platform through SDN network. The IoT gateway and hosts are built in mininet topology. An add on of mininet dashboard is used in sFlow-RT so they can communicate with each other and provide the network flow statistics. In further sub-sections, the installation and working of mininet, sFlow-RT and RYU controller will be done.

**Mininet** Mininet is installed as it is the most utilized platform for SDN topology building. In mininet, the desired topology with multiple hosts and controllers can be created. Currently, in this scenario, four hosts and three switches are considered for flat tree topology. As per further requirements, these switches and hosts count can be increased as mininet is scalable and easy to configure. In Fig. 3, the topology setup is shown with four hosts and three Openflow-enabled switches. One host will be used as IoT gateway that has either MQTT or CoAP protocol based on application domain of IoT. These protocols are currently used in research [9] to send data to ThingsBoard.

**sFlow-RT** The sFlow-RT setup is done, and an add on of mininet dashboard is done in it. Integration of sFlow with mininet such that topology created in mininet can be viewed and monitored in sFlow-RT. sFlow-RT provides GUI port statistics which helps in analysing the flow traffics within the topology created in mininet showed in Fig. 4.

sFlow is an open-source platform for network-related statistical and real-time-based traffic sampling technology for monitoring traffic. It has better GUI to see traffic pattern and network performance. sFlow will be needed in Fig. 1 test bed as later in the project, DDoS attack detection is required for which sFlow would be helpful.

### 4.2.1 RYU Controller

Ryu controller is most widely used controller by researchers to design SDN network [10]. RYU controller is used in this test bed as it is an open-source SDN controller. The script for controller written in Python code supports Openflow switches and many other protocols. It also supports compatibility in creation of well-defined API.

As RYU controller is compatible with OpenFlow switches, Hewlett Packard, IBM and NEC are tested and certified with Ryu controller. It also supports the OpenFlow

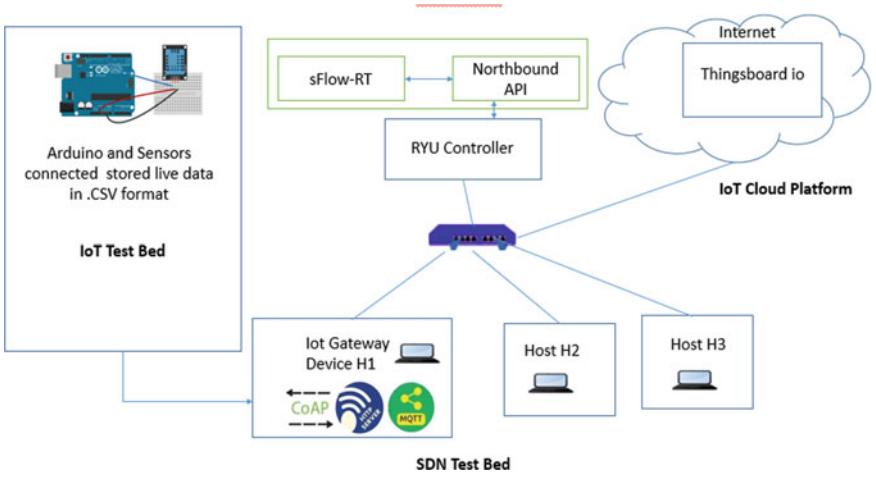



**Fig. 1** IoT-SDN test bed

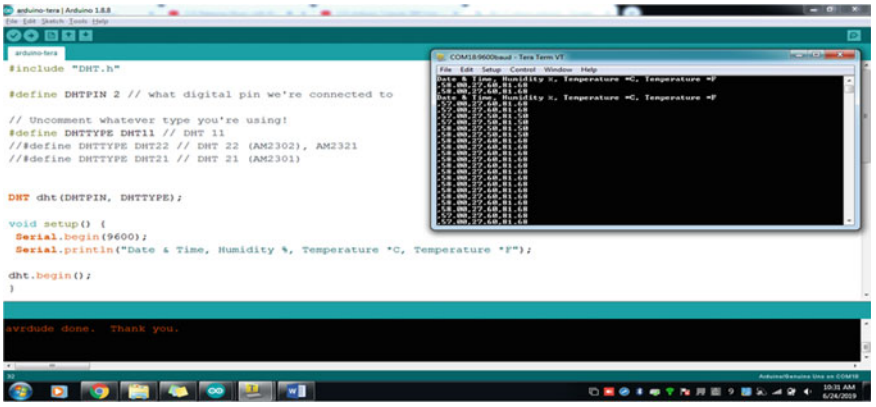



**Fig. 2** Arduino code execution

protocol up to the latest version 1.5 [2]. Ryu controller keeps track of all new and old flow in the network thus monitors the switches connected to it. The Ryu controller is installed and initiated in Fig. 5.

## *4.3 Thingsboard Cloud Platform*

ThingsBoard is based on the Java eight platform. It provides the user with a interface to register and manage devices connected. The credentials can be user-defined strings

```
Connecting to remote controller at 127.0.0.1:6653
*** Creating network
*** Adding controller
*** Adding hosts:
h1 h2 h3 h4
*** Adding switches:
s1 s2 s3
*** Adding links:
(h1, s2) (h2, s2) (h3, s3) (h4, s3) (s1, s2) (s1, s3)
*** Configuring hosts
h1 h2 h3 h4
*** Starting controller
c1
*** Starting 3 switches
s1 s2 s3 ...
*** Enabling sFlow:
s1 s2 s3
*** Sending topology
*** Starting CLI:
```

**Fig. 3** Mininet topology

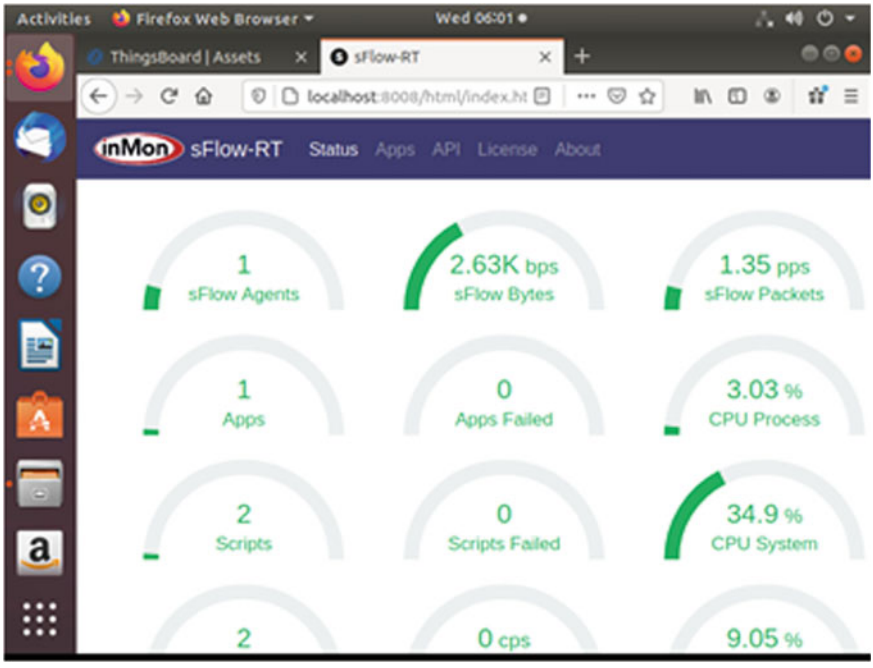



**Fig. 4** sFlow GUI

```
\$ ryu-manager l4_switch.py ryu.app.ofctl_rest
loading app l4_switch.py
loading app ryu.app.ofctl_rest
loading app ryu.controller.ofp_handler
instantiating app None of DPSet
creating context dpset
creating context wsgi
instantiating app l4_switch.py of L4Switch13
instantiating app ryu.controller.ofp_handler of OFPHandler
instantiating app ryu.app.ofctl_rest of RestStatsApi
(19185) wsgi starting up on http://0.0.0.0:8080
```

**Fig. 5** Ryu controller

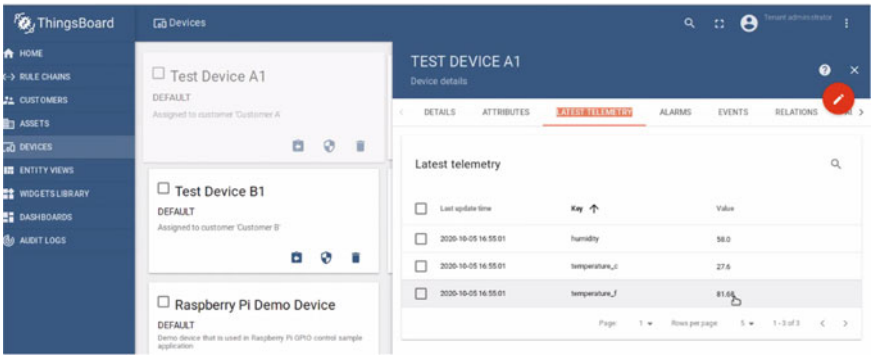



**Fig. 6** ThingsBoard result

called access tokens or X.509 certificates. It stores data received from devices as telemetries of data associated to specific device [11].

The ThingsBoard is cloud IoT platform configured to upload data from IoT environment Fig. 6. The unique key value provided by this platform is given along with protocol for uploading the data. At the end of execution of whole test bed in dashboard, the view of Iot data is visible. The setup required to set up the test bed for SDN and IoT integration is summarized in Fig. 7.

## 5 DDoS Attack and Mitigation Simulation

There are possibilities that in normal network, attacks like denial of service or distributed denial of service attack can take place. In such cases, the attack detection and mitigation in traditional network are time consuming also causes loss of data. When such unreliable network is used for sending time-sensitive data of Iot domain security is great concern. In the proposed architecture with SDN and Iot integration, such attacks are more efficiently detected and mitigated. In current test bed for

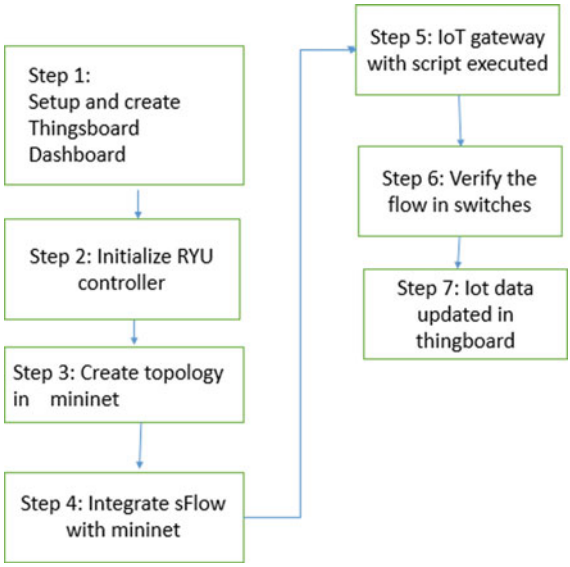



**Fig. 7** Summarized steps of test bed setup

```
containernet> h1 hping3  -S -p 80 -c 10000  h4
HPING 10.0.0.4 (h1-eth0 10.0.0.4): S set, 40 headers + 0 data bytes
len=40 ip=10.0.0.4 ttl=64 DF id=14449 sport=80 flags=RA seq=0 win=0 rtt=19.7 ms
len=40 ip=10.0.0.4 ttl=64 DF id=14605 sport=80 flags=RA seq=1 win=0 rtt=11.6 ms
len=40 ip=10.0.0.4 ttl=64 DF id=14659 sport=80 flags=RA seq=2 win=0 rtt=11.4 ms
len=40 ip=10.0.0.4 ttl=64 DF id=14717 sport=80 flags=RA seq=3 win=0 rtt=11.3 ms
len=40 ip=10.0.0.4 ttl=64 DF id=14948 sport=80 flags=RA seq=4 win=0 rtt=11.2 ms
len=40 ip=10.0.0.4 ttl=64 DF id=14954 sport=80 flags=RA seq=5 win=0 rtt=15.0 ms
len=40 ip=10.0.0.4 ttl=64 DF id=15018 sport=80 flags=RA seq=6 win=0 rtt=18.5 ms
len=40 ip=10.0.0.4 ttl=64 DF id=15150 sport=80 flags=RA seq=7 win=0 rtt=10.4 ms
len=40 ip=10.0.0.4 ttl=64 DF id=15291 sport=80 flags=RA seq=8 win=0 rtt=34.3 ms
len=40 ip=10.0.0.4 ttl=64 DF id=15481 sport=80 flags=RA seq=9 win=0 rtt=19.7 ms
```

**Fig. 8** TCP attack

```
~/mimi $ env "RTPROP=-Dscript.file=$PWD/milestone1/sflow/ddos.js" sflow-rt/start.sh
2020-10-05T16:46:33+05:30 INFO: Starting sFlow-RT 3.0-1519
2020-10-05T16:46:34+05:30 INFO: Version check, running latest
2020-10-05T16:46:35+05:30 INFO: Listening, sFlow port 6343
2020-10-05T16:46:35+05:30 INFO: Listening, HTTP port 8008
2020-10-05T16:46:35+05:30 INFO: /home/suresh/mimi/milestone1/sflow/ddos.js started
2020-10-05T16:46:35+05:30 INFO: app/mininet-dashboard/scripts/metrics.js started
2020-10-05T17:00:29+05:30 INFO: TCP SYN Attack
2020-10-05T17:00:29+05:30 INFO: blocking 10.0.0.1,10.0.0.4,80,000000010
```

**Fig. 9** sFlow attack blocked

simulation of DDoS attack, the traffic can be created using TCP, ICMP or UDP packets. In Fig. 8, TCP packet is flooded in the network between host h1 and h4.

Currently, RYU controller is monitoring the network for adding flow rules and in parallel sFlow-RT is keeping watch over network. In Fig. 9, the API on Ryu controller has detected the TCP flood packets and added the new flow rule to block. The log

```
2020-10-05T16:46:35+05:30 INFO: app/mininet-dashboard/scripts/metrics.js started
2020-10-05T17:00:29+05:30 INFO: TCP SYN Attack
2020-10-05T17:00:29+05:30 INFO: blocking 10.0.0.1,10.0.0.4,80,000000010
2020-10-05T17:02:10+05:30 INFO: unblocking 10.0.0.1,10.0.0.4,80,000000010
```

**Fig. 10** sFlow unblocked

```
sudo ovs-ofctl -O OpenFlow12 dump-flows s1
OFPST_FLOW reply (OF1.2) (xid=0x2):
 cookie=0x0, duration=765.496s, table=0, n_packets=618, n_bytes=49614, priority=0 actions=CONTROLLER:65535
sudo ovs-ofctl -O OpenFlow12 dump-flows s3
OFPST_FLOW reply (OF1.2) (xid=0x2):
 cookie=0x0, duration=766.673s, table=0, n_packets=400, n_bytes=33411, priority=0 actions=CONTROLLER:65535
sudo ovs-ofctl -O OpenFlow12 dump-flows s2
OFPST_FLOW reply (OF1.2) (xid=0x2):
 cookie=0x0, duration=11.285s, table=0, n_packets=1108, n_bytes=46536, priority=4000,udp,in_port=1,nw_dst=10.
0.0.3,tp_src=53 actions=drop
 cookie=0x0, duration=855.173s, table=0, n_packets=629, n_bytes=50802, priority=0 actions=CONTROLLER:65535
```

**Fig. 11** Ryu manager

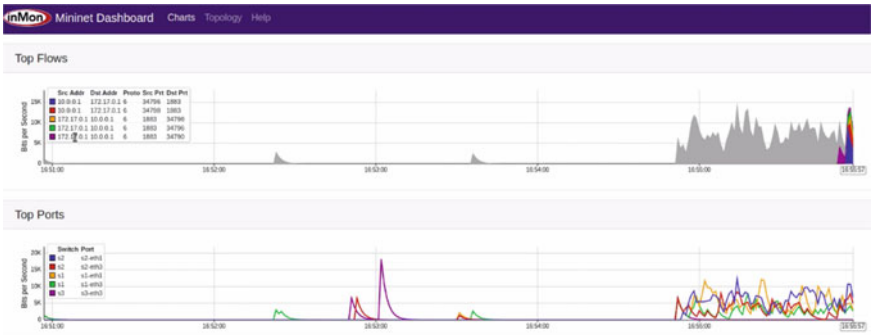



**Fig. 12** DDoS detection in sFlow

record of sFlow shows that TCP SYN attack was detected, and it was blocked. The controller flow rule took action of blocking the switch flow with DROP as action.

sFlow keeps watch on network after blocking it checks every 10 s whether attack has stopped. Once the attack has stopped and no more packets are flooded in network, then the rule is unblocked and normal traffic flow is allowed between these hosts. In Fig. 10, it can be seen that in sFlow-RT log the unblock request is handled.

Similarly, UDP attack can be blocked as well. In Fig. 11, RYU manager it can be seen that both TCP and UDP attacks were taken place, and later, they were blocked by the API created.

In Fig. 12, it can be seen that in sFlow-RT, there are peaks of network traffic due to DDoS attack. In Fig. 13, the peaks are reduced and DDoS mitigatio is done. sFlow detects the DDoS attack and informs the controller through Northbound REST API which blocks the flow of traffic. Later, after a time interval, it reconfirms if attack not taking place then acceptance flow rule added and its unblocked as well.

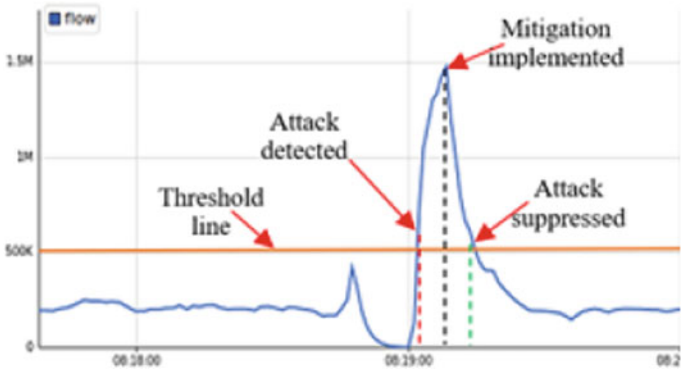



**Fig. 13** DDoS detection and mitigation

## 6 Conclusion

In the test bed created for IoT-SDN-cloud integration, it enables the user to send end to end starting from collection of raw data from live IoT environment to cloud using SDN as backbone network. The advantages of SDN and IoT integration are recognized in many domains like smart grid settings, smart homes or smart transportation. SDN-IoT integration is also provided security in IoT, because security mechanisms can be implemented easily. The sFlow-RT can monitor the attacks and inform Northbound REST API over RYU controller. The Northbound API is programmable to mitigate these attacks. The test bed provides scope for many researchers to do further research or analysis on different network parameters like comparative study between different SDN controllers, IoT protocols or scope of API for resolving different security issues and load balancing.

## References

1. Cherian M, Chatterjjee M (2018) Survey of security threats in IoT and emerging countermeasures. In: International symposium on security in computing and communication SSCC 2018: security in computing and communications, pp 591–604
2. Bull P (2016) Flow based security for IoT devices using an SDN gateway. In: IEEE 4th international conference on future internet of things and cloud (FiCloud), Austria, 2016, pp 157–163
3. Flauzac O (2015) SDN based architecture for IoT and improvement of the security. In: Proceedings of the IEEE 29th international conference on advanced information networking and applications workshops (WAINA), South Korea, 2015, pp 688–693
4. Gonzalez C (2016) A novel distributed SDN-secured architecture for the IoT. In: Proceedings of the IEEE international conference on distributed computing in sensor systems (DCOSS), Washington, USA, 2016, pp 244–249
5. Jararweh Y, Al-Ayyoub M, Darabseh A, Benkhelifa E, Vouk M, Andy R (2015) SDIoT: a software defined based internet of things framework. J Ambient Intell Hum Comput

6. Bhunia SS, Gurusamy M (2017) Dynamic attack detection and mitigation in IoT using SDN. In: 27th international telecommunication networks and applications conference (ITNAC). IEEE
7. Zheng S (2019) Research on SDN-based IoT security architecture model. In: 8th joint international information technology and artificial intelligence conference (ITAIC 2019)
8. Huang H, Zhu J, Zhang L (2014) An SDN based management framework for IoT devices. In: IEEE conference on computer communications (INFOCOM). IEEE, 2015, pp 208–216; Irish signals & systems conference 2014 and 2014 China–Ireland international conference on information and communications technologies (ISSC 2014/CIICT 2014)
9. Theodorou T, Mamatas L (2017) CORAL-SDN: a software-defined networking solution for the internet of things. In: 2017 IEEE conference on network function virtualization and software defined networks (NFV-SDN)
10. Bedhief I, Kassar M, Aguili T (2018) From evaluating to enabling sdn for the internet of things. In: 2018 IEEE/ACS 15th international conference on computer systems and applications (AICCSA)
11. Ismail AA, Hamza HS, Kotb AM (2018) Performance evaluation of open source IoT platforms. In: 2018 IEEE global conference on internet of things (GCIoT)
12. Asadollahi S, Goswami B (2018) Ryu controller's scalability experiment on software defined networks. In: 2018 IEEE international conference on current trends in advanced computing (ICCTAC)
13. De Paolis LT, De Luca V, Paiano R (2018) Sensor data collection and analytics with ThingsBoard and Spark Streaming. In: 2018 IEEE workshop on environmental, energy, and structural monitoring systems (EESMS)
14. Lawal BH, Nuray AT (2019) Real-time detection and mitigation of distributed denial of service (DDoS) attacks in software defined networking (SDN). IEEE
15. Sambandam N, Hussein M, Siddiqi N, Lung C-H (2018) Network security for IoT using SDN: timely DDoS detection. IEEE
16. Pande B, Bhagat G, Priya S (2018) Detection and mitigation of DDoS in SDN. In: Eleventh international conference on contemporary computing (IC3), 2–4 Aug 2018
17. Sun W, Li Y, Guan S (2019) An improved method of DDoS attack detection for controller of SDN. In: IEEE conference on computer communications (INFOCOM), 2nd international conference on computer and communication engineering technology—CCET. IEEE
18. Pasumponpandian A, Smys S (2019) DDOS attack detection in telecommunication network using machine learning. J Ubiquitous Comput Commun Technol (UCCT)
19. Chen JIZ, Smys S (2020) Social multimedia security and suspicious activity detection in SDN using hybrid deep learning technique. J Inf Technol Digit World (2020)

# Low Rate Multi-vector DDoS Attack Detection Using Information Gain Based Feature Selection


**R. R. Rejimol Robinson and Ciza Thomas**

**Abstract** The number of connected devices is exponentially growing in the world today and they need to work without having any interruption. This scenario is very challenging to cybersecurity and needs proper attention of network administrators, service providers, and users. Implementing security frameworks in this scenario is very difficult because attackers are using very sophisticated easy to operate weapons to launch huge attacks such as Distributed Denial of Service. Intelligently detecting and mitigating the attacks in the network requires the use of machine learning algorithms. This work proposes a strategic way involving feature selection based machine learning for the detection of stealthy attacks. The detection system works by performing ınformation gain-based feature selection as a preprocessing step. This ensures case-based preprocessing of each attack vector present in the traffic and is proved to be effective empirically. The proposed method has been tested using two supervised machine learning classification algorithms, namely Random forest and J48. The evaluation results show that the Random forest algorithm gives a satisfactory True Positive rate of 99.6% in detecting stealthy layer 7 attacks. The overall accuracy obtained is 99.81%. This approach causes the algorithms to exhibit improved performance while doing classification.

**Keywords** Machine learning · Feature selection · Low rate attacks · Information gain · Stealthy attacks · Network security

## 1 Introduction

The digitally connected modern world demands uninterrupted connections, even the disruptions are unavoidable. Distributed Denial of Service (DDoS) attack is one such

R. R. Rejimol Robinson (✉)
SCT College of Engineering, Thiruvananthapuram, India
e-mail: rejibz@sctce.ac.in

C. Thomas
Directorate of Technical Education, Thiruvananthapuram, Kerala, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_53

annoyance that makes an online service unavailable. Its impact is devastating unless it is detected and mitigated properly. Formally, define the DDoS as a cyber-attack launched with overwhelming traffic from multiple sources to make the target machine (server) or other network resources unavailable to its intended users temporarily or indefinitely. So it is impossible to stop the attack by simply blocking a single source of the attack. It is the responsibility of network administrator to monitor and supervise their network and guarantee the proper functioning of the network.

There are several strategies for launching an attack and the most prevalent used in these days are zomby attacks or otherwise known as botnet attack. The attack is carried out by a handler that infects vulnerable hosts and recruits them for their purpose. These machines are called zombies on the ınternet and under the control of the handler, zombies are directed to launch an attack by bombarding false packets towards a target to limit its performance or crash it.

Instead of depending on traditional methods of DDoS detection based on Firewalls and Intrusion Detection System (IDS), it is desirable to switch on to machine learning and deep learning based detection strategies. The detection and mitigation system needs to be more agile and intelligent as the attacks are getting more and more sophisticated. But the performance of algorithms is affected due to massive and high-dimensional data received in real-time. Worldwide Infrastructure Security Report by Netscout declared that hackers entered into the era of terabit attacks. According to them, 1.7 TBPS DDoS was recorded in the year 2020 [1]. It indicates the fact that DDoS attack has continued to increase both in size and sophistication.

The stealthiness of DDoS attacks is also getting higher such that multi-vector, slow, and low rate attacks are common. Combination of volumetric and protocol attacks aiming at different layers of network such as layer 7, transport, and network layer are very common nowadays. Multi-vector attacks are also known as polymorphic cyber attacks. They use two or more methods of infiltration. They are launched for a variety of reasons. Some attacks aim to steel sensitive data which is later handed over to a third party merely for monetary benefit and some attacks to take down the network. These are automated attacks and they dynamically change parameters and vectors in response to the defense mechanisms. The Corero Network Security Inc., explained in their blog that sometimes the multi-vector attacks layer different vector types and sometimes they vary the attack vector itself to evade detection. When it is continually modified, it becomes much more difficult to mitigate them. They often start with one vector, such as a simple UDP flood and, if unsuccessful, they try a second technique such as a DNS flood [2]. Hence it is very difficult to detect modern DDoS attacks due to their varying features and different points of entry to infiltrate the network. In this scenario, it is highly essential to do some intelligent preprocessing steps before the machine learning algorithms are applied for detection.

Network traffic data processing comes in two flavors namely, packet-based and flow-based approaches. A flow can be formally defined as a unidirectional stream of IP packets that share a set of common properties; the IP-protocol, source, destination IP addresses, source, and destination ports used. It is often desirable to analyze data inflows rather than packets since it greatly reduces the complexity of data. It is logical

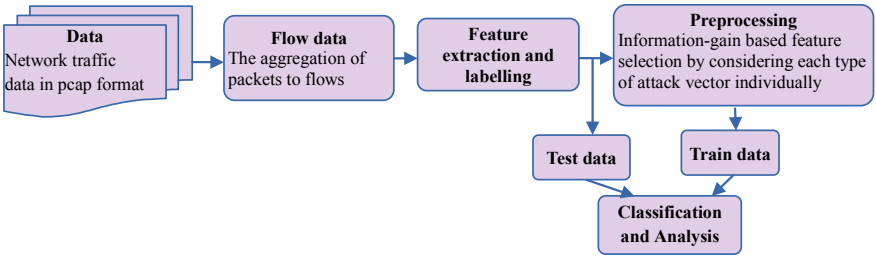



**Fig. 1** Workflow diagram of the proposed method

to consider a connection as benign or anomalous rather than packets and hence flow-based analysis is the efficient and fast way of detection. Hence network traffic in the form of flows is used in the proposed work.

The proposed work, mainly concentrating on modern stealthier attacks especially multi-vector stealthy attacks. The workflow diagram showing the conceptualization of the proposed work is given in Fig. 1. The number of benign packets passing through the network per unit time at a particular point is enormous compared to several attack instances. So DDoS attack detection is regarded as an imbalanced dataset problem. Hence it is required to make the features more and more bright to make detection easier. So feature selection is considered an essential preprocessing step in this work. According to Jain et al, feature selection is an effective way of dimensionality reduction and can reduce the complexity of attack detection and thereby increase the detection accuracy [3]. To employ an intelligent strategy initially select the features, and the stealthier attacks. Stealthiness is considered as the resemblance of attack vectors to benign traffic. Different attacking vectors have different prominent features. Hence case-based feature selection is proposed, where each attack vector is treated separately to select features. Making features very prominent is crucial, as it is an imbalanced dataset problem. Feature selection algorithm which can project the maximum information regarding minority attack samples is recommended. So ınformation gain-based feature selection is proposed for doing preprocessing. It is one of the popular feature selection methods due to its computational efficiency and simplicity. This method selects a feature that has high relevance to the output class and also has the highest occurrence rather than simply performing the dimension reduction.

There are two algorithms selected namely J48 and Random forest for doing the performance analysis. The parameters of the algorithms are set in their default settings and it is proposed to evaluate the algorithms with and without doing the information gain-based feature selection on the training data. The important factor of the proposed method is that the classification is being done using a supplied test set rather than doing cross-validation.

The contributions of this work involve:

1. Stealthier multi-vector attack detection with Information gain-based feature selection to get an optimal set of features based on its distinguishing property related to each attacking vector.
2. Enhancing the performance of machine learning algorithms in detecting stealthier attacks.
3. Comparison of methodology with results obtained in other related literature.

## *1.1 Stealthy Attack Variants*

(a) *Low and slow attacks*

These types of attacks use a low volume of data and operate very slowly. Designed to send small amounts of data across multiple connections to keep ports on a targeted server open as long as possible, these tools continue to utilize server resources until a targeted server is unable to maintain additional connections. Uniquely, low and slow attacks may be effective even when not using a distributed system such as a botnet and are commonly used by a single machine.

(b) *Slowloris*

Apart from being a slow-moving primate, Slowloris is an application designed to investigate a low and slow attack on a targeted server. The elegance of Slowloris is the limited amount of resources it needs to consume in order to create a damaging effect.

(c) *TCP flooding*

The attacker exploits the three-way handshake of TCP communication. The attacker sends SYN packets to the server pretending to establish a TCP connection, the server sends SYN-ACK packet back to the client and keeps a port open to receive ACK from the client, but the attacker never sends a final ACK to the server. So the attacker keeps on sending the SYN packets to the server and the server keeps opening a port temporarily for a specific time. The server stops working and responding to legitimate clients after all the ports are utilized.

(d) *UDP attack*

In a UDP attack, the attacker sends a bogus UDP packets to the server using a random port. The server is actually looking for the application on that port. If the service was not running on that port, the server replies with ICMP unreachable message to the attacker. The attacker continuously sends UDP packets and the server also replying ICMP unreachable message back to clients, that will lead to maximum resource consumption of the victim and the network as well. Eventually, the server can not respond to its legitimate user.

The rest of the sections in this article is organized as follows. Section 2 deals with the detailing of the literature survey. Section 3 describes the proposed case-based

feature selection methodology. Section 4 analyze the results and conclusion of the work is presented in Sect. 5.

## 2 Literature Survey

Due to the proliferation of data to be handled in this new digital world and the high dimensionality of collected data, feature selection is considered as one of the most important factors in determining the efficiency of detection systems. But to have an efficient model, the redundant and less sensitive features need to be dropped.

Combinations of feature selection methods with other optimization techniques are employed in almost all the application domains to improve the detection accuracy of machine learning algorithms. The work of Chuang et al, achieved comparable accuracy when K-nearest neighbor (KNN) with the leave-one-out cross-validation (LOOCV) is used for classifying eleven different gene expression data set. The preprocessing technique employed to bring out this performance is by doing hybrid feature selection methods involving correlation-based feature selection (CFS) and the Taguchi-genetic algorithm (TGA) [4].

According to the work of Gunal et al., a hybrid of both filter and wrapper feature selection steps is being proposed to analyze the redundancy or relevancy of the text features. The experiment done in this work proved the effectiveness of selecting features using different methods rather than stick on to a single method. The combination of the features selected has a profound impact on text classification [5].

The work of Wang et al, mainly deals with DDoS detection based on feature selection involving SU genetic algorithm. The number of features of the NSL-KDD dataset, reduced to 17 from 41 and the machine learning algorithms such as J48 and Random Forest yields 99.8% of detection accuracy. [6] Singh et al, in their work, considers Naive Bayes as the classifier and use information gain-based feature selection and attained 99.5% accuracy in detecting DDoS traffic present in CAIDA 2007 dataset. They have analyzed the results on packet-based data and the features selected are SYN value, ACK value, and Time To Live (TTL) [7].

According to the work of Osanaiye et al., the Ensemble-based Multi-Filter Feature Selection (EMFFS) method is the amalgamation of Information Gain (IG), Gain Ratio, Chi-squared, and relief is used to select important features. The experiments are done on the NSL-KDD dataset and the accuracy is found to be 99.6% using the J48 algorithm. The number of features employed in this method are 13 [8].

In the work of Kamarudin et al, a hybrid feature selection model combines the strengths of the filter and the wrapper feature selection procedure. This hybrid solution selects the optimal set of features in detecting attacks. Correlation feature selection (CFS) together with three different search techniques known as best-first, greedy step-wise, and genetic algorithm are used. The wrapper-based subset evaluation uses a random forest classifier to evaluate each of the features that were first selected

by the filter method. Tested on KDD99 and DARPA 1999 dataset with ten-fold cross-validation in a supervised environment and yields satisfactory results [9].

Lima et al., proposed a smart detection method using machine learning, and this work is designed to detect both high and low volume DDoS attacks. The preprocessing method such as Recursive Feature Elimination with Cross-Validation (RFECV) is used in this work. The datasets CIC-DoS, CICIDS 2017, CSE-CIC-IDS2018, and the ISCXIDS2012 Dataset are mainly employed in this work to evaluate the model [10].

Gu et al., proposed a semi-supervised weighted k-means detection method. A Hadoop-based hybrid feature selection method is used to find the most effective feature set. Then a semi-supervised weighted k-means method using hybrid feature selection algorithm (SKM-HFS) is employed to achieve better performance. The datasets used are DARPA DDoS dataset, CAIDA DDoS attack 2007 dataset, CICIDS DDoS attack 2017 dataset, and real-world dataset [11].

In the work of Wu et al., employed a hybrid feature selection method to detect network anomalies [12]. Wang et al., proposed a combination of sequential feature selection with dynamic MLP to select the optimal features during the training phase and designed a feedback mechanism to reconstruct the detector when perceiving considerable detection errors. The work is mainly done on the NSL-KDD dataset, ISOT, and ISCX [13].

## 3 Proposed Methodology

During the network traffic data in PCAP format utilize the UDP, ICMP, and TCP flooding attack packet which traces the CAIDA dataset to form multi-vector attack. TC Preplay tool is mainly employed to generate such traffic. TCPreplay is a free Open Source utility suite for editing and replaying previously captured network traffic. It is designed to replay malicious traffic patterns to the Intrusion Detection/Prevention System. All the remaining experiments are done using Python Pandas and WEKA. Pandas is a software library written for the Python programming language for data manipulation and analysis. WEKA, a data mining software, and an open-source software facilitates data preprocessing and implementation of several machine learning algorithms [17].

There are two important phases, one of which is the strategic feature selection phase and the other one deals with the evaluation of machine learning algorithms. Before going to the actual preprocessing step, the data in PCAP format must be processed to form flows and along with that, features are also extracted. Data is the real matter of concern while evaluating the method. CAIDA (Center for Applied Internet Data Analysis) DDoS Attack 2007 dataset is selected to form the sufficient dataset for multi-vector attacks. These traces consist of TCP, UDP, and ICMP flooding instances [15]. To analyze low and slow attacks, Canadian Institute of cybersecurity, CICIDS 2017 dataset has been selected. This dataset contains realistic background traffic and up-to-date attacks [16]. CAIDA data comes in PCAP format only. To implement a

feature extraction module the above process is handled. It extracts features of each flow related to the packet trace namely: Average_Packet_Size, Number_of_Packets, Time_Interval_Variance, Packet size_Variance, Number_of_Bytes, Packet_Rate, and Bit_rate are the seven extracted features [18]. The feature vector formed in this process is represented as $X = (x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7)$, where $x_i$ is the $i$th feature extracted from the traffic flow, $x_7$ is label and $X$ represents the feature vector by the feature extraction module. The CICIDS 2017 dataset comes in CSV format, a collection of flow instances with a total of 79 attributes including the label.

The prepared data is represented in CSV format and is ready to give as input to the next phase. The steps involved in the case-based feature selection using information gain is depicted in Fig. 2. The supervised learning algorithms are evaluated to prepare training and test dataset and they are set as 70% of train and 30% test. Then the training data is split into subsets of training data which comprises any one kind of attack vector and normal data. Then the CfsSubsetEval and information gain based feature selection methods are applied to these subsets. The features at the top positions in the rank list produced by these algorithms are selected.

For having a comparison with other feature selection methods, employ the CfsSubsetEval algorithm available in WEKA. This is a correlation-based feature selection method and is based on the logic such that, relevant feature subsets contain features

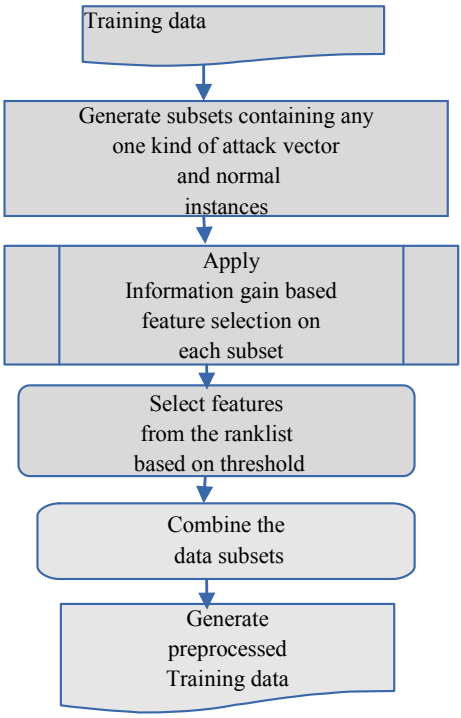



**Fig. 2** Flowchart of preprocessing steps involved

highly correlated with the classification, but uncorrelated to each other. The Information Gain (IG) based feature selection is suitable for multi-vector attack feature detection, as it can select the features which contain more relevance which in turn makes the detection easier. It quantifies information based on the contribution of the presence or absence of a feature in making the correct classification decision for any class and is computed according to Eq. (1). The higher the value of mutual information between classes $C$ and feature $f$, the higher the relevance of feature $f$ in classes $C$.

$$IG(C, f) = H(C) - H(C|f) \tag{1}$$

where, $H(C) = -\sum_{c \in C} p(c) \log p(c)$ the entropy of the class $C$ and $H(C|f) = -\sum_{c \in C} p(c|f) \log p(c|f))$ is the conditional entropy of class given feature $f$. The minimum value of $IG(C, f)$ means that $H(C|f) = 1$. That is class $C$ and feature $f$ are not at all related. That is the most distinguishing feature will be the one that is related to a particular class. Due to the feature interactions among the instances, it is not possible to have the ideal case [19].

The next step is to build the new training and test dataset with these selected features. A threshold is set such that those features having information gain value greater than the threshold value are selected as best features. Then new train data is created by joining the subsets of data and it is shuffled to evenly distribute the instances. The next phase of the proposed methodology is the evaluation of machine learning algorithms. J48, the WEKA implementation of decision tree and Random forest algorithms are selected for evaluation. Our earlier work in which the ranking of ten machine learning algorithms has been done. J48 and Random Forest are the algorithms placed in the topmost position in detecting DDoS attacks [20].

A Decision Tree (DT) is one of the well-known supervised classification algorithms in which a tree is generated which acts as a multistage decision system. The concept is based on the measure of the variance of data which demonstrates the presence of different categories of data. A feature vector is assigned with a class label through a sequence of Yes/No decisions along a path of nodes of a DT. The most important factor of splitting criteria for a particular node is to decrease the entropy such that homogeneous vectors can be brought under one particular node since entropy is the measure of impurity [21].

Random forest is an ensemble classification method that combines a collection of classifiers (i.e., decision trees) to make a "forest". Each of the decision trees is generated by using a random selection of attributes at each node to determine the split [22].

## 4 Result and Discussion

The machine learning algorithms are evaluated based on the True Positive rate (TP) of detecting attack vectors. Among the variety of evaluation metrics such as precision,

recall, F-measure, etc, TP rate is selected to deal with highly skewed data of network traffic. The true positive rate or sensitivity is calculated as given in Eq. (2). TPR is the probability that an actual positive will test positive. TP is the number of instances predicted as positive and TP+FN is the total number of positive samples present in the dataset, where FN is the positive samples misclassified. Misclassification of attacks is costlier than misclassifying benign traffic, so more importance is given to the TP rate of attacks rather than normal traffic. The TP rate comparison is given in Tables 1 and 2.

$$\text{TP\_rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

But when evaluating the J48 and Random Forest with this dataset, the detection rate of SlowHTTPtest is considerably very low compared to other attack vectors. The proportionality of the Hulk attack is very high compared to other attacks in the dataset. 33.3% Hulk attack is there while only 0.7% SlowHTTPtest, 0.8% slowloris, and 1.48% GoldenEye in the dataset. The considerable distributional overlap is there between SlowHTTPtest and slowloris attacks. So the TP rate obtained for SlowHTTPtest is very low compared to other attack variants. The features of the GoldenEye attack are very prominent, so the TP rate obtained for this attack is competitively good without any preprocessing. The proportionality of attacks in the CAIDA 2007 dataset is also very low and is 2% only.

The result of the preprocessing step is very important to be analyzed. The CICIDS dataset comes with a total of 79 features and CAIDA 2007 dataset contains 8 features.

**Table 1** Table showing the results of CAIDA 2007 dataset

| TP rate | | | | | | |
|---|---|---|---|---|---|---|
| Predicted | CFS | | Information gain | | Without preprocessing | |
| | J48 | Random forest | J48 | Random forest | J48 | Random forest |
| Normal | 0.98 | 0.994 | 0.993 | 0.996 | 0.981 | 0.992 |
| Attack | 0.943 | 0.989 | 0.965 | 0.995 | 0.995 | 0.99 |

**Table 2** Table showing the results of CICIDS 2017 dataset

| TP rate | | | | | | |
|---|---|---|---|---|---|---|
| Predicted | CFS | | Information gain | | Without preprocessing | |
| | J48 | Random forest | J48 | Random forest | J48 | Random forest |
| Benign | 0.999 | 0.999 | 1 | 1 | 0.995 | 0.996 |
| Slowloris | 0.991 | 0.993 | 0.994 | 0.994 | 0.991 | 0.993 |
| Slowhttptest | 0.875 | 0.874 | 0.985 | 0.988 | 0.856 | 0.865 |
| Hulk | 1 | 1 | 1 | 1 | 1 | 1 |
| Goldeneye | 0.989 | 0.989 | 0.996 | 0.997 | 0.989 | 0.989 |

**Table 3** Comparison of accuracy obtained

| Dataset algorithm | CAIDA2007 | CICIDS 2017 (%) |
|---|---|---|
| Random forest | 99.56% | 99.81 |
| J48 | 98.21% | 99.77 |

The number of features selected is greatly reduced especially in the case of CICIDS 2017 dataset. CFSsubsetEval algorithm selects only 6 features for CICIDS 2017 dataset as a whole namely: Destination_Port, Total Length_of_Packet, Bwd_Packets, Init_Win_bytes_forward, Init_Win_bytes backward, and Idle_Max. Similarly, only two features are selected for CAIDA 2007 dataset namely, Time_Interval_Variance and Number_of_Packets. Competitively the number of selected features is more in the case of case-based feature selection. The prominent features of each attack vector are selected separately. It makes each attack more distinguishable even though their proportionality is very low. The bottom line is that a combination of the features selected by information gain based feature selection methods is more effective than the features selected by the CfsSubsetEval. The performance of the machine learning algorithm depends on so many factors among which feature selection is very important. The other important factor is data proportionality. The results show a considerable improvement in detection rate even without any oversampling or synthetic sampling steps, which are the normal preprocessing steps for an imbalanced dataset problem. Effective discriminating power is increased when all the individually selected features together were concatenated.

For having a comparison with the literature published very recently, compute the detection accuracy which is given in Table 3. The literature selected for comparison is briefed in Table 4. The literature which explains the work on modern stealthy attack datasets such as CICIDS 2017 and CAIDA 2007 were considered. This comparison

**Table 4** Summary of literature used for comparison

| Reference | Feature selection | Model | Dataset | Accuracy (%) |
|---|---|---|---|---|
| Singh et al. [7] | Information gain | Naive Bayes | CAIDA 2007 and CAIDA anonymous trace 2015 | 99.5 |
| Lima et al. [10] | NA | Random forest | CIC-DoS, CICIDS2017 and CSE-CIC-IDS2018, | 98.6 |
| Gu et al. [11] | Hadoop based hybrid feature selection | Semi-supervised K-means algorithm | DARPA DDoS dataset, CAIDA DDoS attack 2007, CICIDS DDoS attack 2017 | 99 |
| Singh et al. [14] | NA | Multi layer perceptron with a genetic algorithm | CAIDA 2007 | 98.04 |

shows improved performance of machine learning algorithms can be achieved using the proposed method.

## 5 Conclusion

This work aims to detect stealthier DDoS attacks such as low rate Layer 7 attacks and multi-vector attacks present in the network traffic. Hence it is proposed to have an intelligent approach by preprocessing each attack vector separately to select the most bright features using Information gain. The method is tested with UDP, TCP, ICMP protocol attacks of the CAIDA dataset and the layer 7 low rate attacks such as slowloris, Slowhttptest, Hulk, and GoldenEye attacks of CICIDS 2017 datasets. Machine learning algorithms selected are Random forest and J48 and they are evaluated mainly based on the TP rate of detecting attack vectors. Much improvement of learning algorithms can be claimed over the sub-optimal performance exhibited while performing the classification in the highly imbalanced network traffic. The true positive rate of SlowHTTPtest attack, without any preprocessing is 0.875 but it is improved to 0.985 according to the proposed method and an average TP rate of 0.993 is achieved in detecting multi-vector attacks when tested with supplied test data. The proposed method works superior to any other works found in other literature examined especially in detecting low rate and multi-vector attacks. The accuracy obtained for the Random forest algorithm on CICIDS 2017 dataset is 99.81%.

The work can be extended to include hybrid feature selection methods and can be tested with other combinations of machine learning algorithms as well. Selecting the more distinguishing feature or make the features more and more prominent is the preprocessing method required to make the attack detection more effective.

## References

1. Distributed denial of service attack threat report by Netscout. https://www.netscout.com/report/
2. Report on modern DDoS attacks. https://www.corero.com/blog/understanding-and-\\stopping-multi-vector-ddos-attacks/
3. Jain A, Zongker D (1997) Feature selection: evaluation, application, and small sample performance. IEEE Trans Pattern Anal Mach Intell 19(2):153–158 (1997)
4. Chuang L-Y, Yang C-H, Wu K-C, Yang C-H (2011) A hybrid feature selection method for DNA micro-array data. Comput Biol Med 41(4):228–237
5. Gunal S (2012) Hybrid feature selection for text classification. Turkish J Electr Eng Comput Sci 20(2):1296–1311
6. Wang C, Yao H, Liu Z (2019) An efficient DDoS detection based on Su-genetic feature selection. In: Cluster Comput 22(1):2505–2515
7. Singh NA, Singh KJ, De T (2016) Distributed denial of service attack detection using Naive Bayes classifier through info gain feature selection. In: Proceedings of the international conference on informatics and analytics, pp 1–9

8. Osanaiye O, Cai H, Choo K-KR, Dehghantanha A, Xu Z, Dlodlo M (2016) Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. EURASIP J Wirel Commun Netw 2016(1):130
9. Kamarudin MH, Maple C, Watson T (2019) Hybrid feature selection technique for intrusion detection system. Int J High Perform Comput Netw 13(2):232–240
10. Lima Filho FSD, Silveira FA, de Medeiros Brito Jr A, Vargas Solar G, Silveira LF (2019) Smart detection: an online approach for DoS/DDoS attack detection using machine learning. Sec Commun Netw
11. Gu Y, Li K, Guo Z, Wang Y (2019) Semi-supervised K-means DDoS detection method using hybrid feature selection algorithm. IEEE Access 7:64351–64365
12. Wu H, Zhang B, Dong S (2015) A hybrid feature selection method for network traffic anomaly detection. J Phys Conf Ser 1395(1):1. IOP Publishing, 2019
13. Wang M, Lu Y, Qin J (2020) A dynamic MLP-based DDoS attack detection method using feature selection and feedback. Comput Sec 88:101645
14. Singh KJ, De T (2017) MlLP-GA based algorithm to detect application layer DDoS attack. J Inform Sec Appl 36:145–153
15. The CAIDA UCSD DDoS attack 2007 dataset. https://www.caida.org/data/passive/ddos-20070804dataset.xml
16. The CİCİDS 2017 dataset. https://www.unb.ca/cic/datasets/ids-2017.html
17. Hall M, Frank E, Holmes G, Pfahringer B, Reute-mann P, Witten IH (2009) The weka data mining software: an update. In: ACM SIGKDD explorations newsletter, vol 11(1), pp 10–18
18. Karimazad R, Faraahi A (2011) An anomaly-based method for DDoS attacks detection using RBF neural networks. In: 2011 international conference on network and electronics engineering, IPCSIT, vol 11
19. Kent JT (1983) Information gain and a general measure of correlation. Biometrika 70(1):163–173
20. Robinson RR, Thomas C (2015) Ranking of machine learning algorithms based on the performance in classifying DDoS attacks. In: 2015 IEEE recent advances in intelligent computational systems (RAICS). IEEE, pp 185–190
21. Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106
22. Liaw A, Wiener M et al (2002) Classification and regression by random forest. R news 2(3):18–22

# A Framework for Monitoring Patient's Vital Signs with Internet of Things and Blockchain Technology


**A. Christy, MD Anto Praveena, L. Suji Helen, and S. Vaithyasubramanian**

**Abstract** In this prevailing situation of automation, private data related to the public needs to be stored for all businesses and transactions. Most often, the private data of the patient's name, mobile number, diseases, laboratory report, and treatment undergoing is breached by intruders. Automation in the healthcare sector can be made in a highly secured and less complicated manner by integrating patient's health records and health insurance agencies by adopting blockchain technology. In this article, a real-time patient health monitoring system has been built and implemented on an IoT platform which can allow healthcare agencies like hospitals and doctors to monitor critical data of the patient in real-time and respond to the needs of the patient accordingly. The proposed system will prove to be beneficial for patients as well as the doctors and nurses and will lead to the implementation of an intelligent system which can automate various overviews being conducted by healthcare agencies adopting blockchain technology.

**Keywords** Smart healthcare · Blockchain · Internet of things · Body area network · Sensors · IoT tools

## 1 Introduction

Internet of things (IoT) was first proposed by Kevin Ashton who had worked at MIT's Auto-ID Center though the overall concept of implementation of a system architecture that is connected to a network and can be likely monitored by the use of smart devices date back almost four decades back when a Coke machine at Carnegie Mellon University was connected to the Internet which allowed the vendors to prepare reports on the drinks. In the present context, the Internet of things is recognized as the next big step in the field of technology alongside artificial intelligence, big

A. Christy (✉) · M. Anto Praveena · L. Suji Helen
Department of CSE, Sathyabama Institute of Science and Technology, Chennai, India

S. Vaithyasubramanian
PG and Research Department of Mathematics, D. G. Vaishnav College, Chennai, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_54

data, blockchain among others, which allows many devices and gadgets traditionally connected through Internet, to interact with each other and share data, which would then allow the user to automate various tasks.

Every IoT application consists of various layers in which the perception layer occupies the top of the system hardware and consists of all the hardware which includes chips, sensors, and actuators which collects data from various nodes and gadgets connected to the Internet, which it then proceeds to send through an underlying networking layer. The networking layer plays the role of an interface between the hardware layer and the IoT appliance. The collected data can be fed on a NoSQL framework which is then analyzed by algorithms and the processed data is visualized on a user interface which allows the user to draw inferences from the data collected from various ends. The user interface is implemented through a computer/smartphone (Windows, iOS, Android) application which makes the whole process more versatile and automated by the usage of embedded systems.

In recent years, there has been a surge in the adoption of new and novel technologies in the field of medical science and treatment to make the medical and healthcare machinery more efficient. In this field, the Internet of things can operate as a natural amendment to the problems that are faced in the field of real-time monitoring of patient's health parameters. The proposed system employs the use of three sensors namely, peripheral pulse oximetry to measure the oxygen content in the blood, LM35 temperature sensor to measure the temperature of the human body, TCRT1000 transmitter to measure the heartbeat and finally a blood pressure sensor that can assess the muscular functions in arteries and veins to calculate the blood pressure which can be displayed in real-time. Using these sensors, aided by a power supply unit and a Raspberry Pi microcontroller, the data can be displayed by processing it on an application that will allow the doctors to assess the patient's data in real-time (Fig. 1).

The Raspberry Pi functions as the heart of the whole system as it connects all the sensors in a single chip which is wearable and is connected to the Internet by a MAC address, with the data completely available on the website which can be accessed and graphed into suitable data for various applications. The implementation of such a

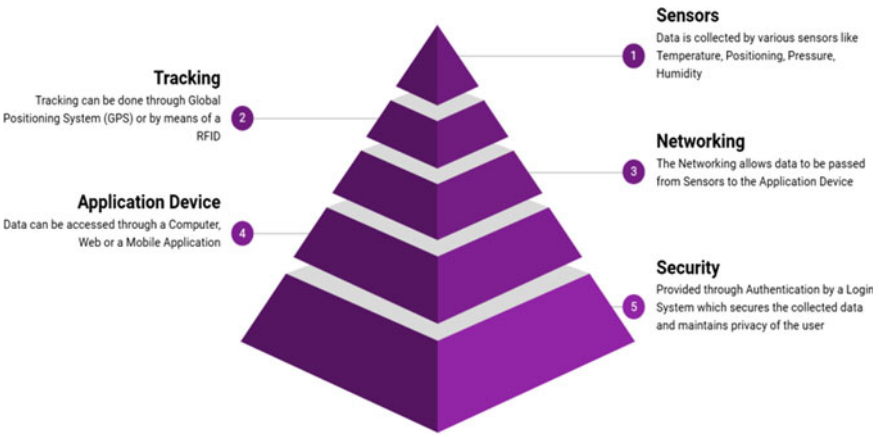



**Fig. 1** Connectivity of IoT

system will be a big step in revolutionizing the automation of healthcare management and hospitals as it allows the concerned agencies to keep track of multiple individuals in a single time by using the Internet of things.

## 2 Motivation

Security and privacy remain a major concern in the Healthcare sector. Privacy deals with granting or revoking accessing rights to private information with others. Privacy determines the actors involved with sharing and managing the information. Consensus with providers of healthcare and regulators can make privacy effective. Security standards such as HIPAA, COBIT, and DISHA are deployed to protect the EHR of the patients. With the availability of big data and IoT, many countries have developed security standards and regulations to protect the patient's EHR.

Data related to a patient is collected from heterogeneous data sources, and often, they are remaining in the form of duplicated, inconsistent, and incompleteness. The bulk of repositories owned by hospitals, healthcare providers, and pharmaceutical companies do not share information among themselves. Security breaches can happen easily some data contain patient's private information such as addresses, mobile numbers, previous history, etc. Currently, most healthcare systems use centralized client–server architectures, in which data is maintained in a centralized server. Intruders can hack the system and manipulate patient's data which is of primary importance.

The industry version 4.0 has brought in revolutions in the form of cloud computing, fog computing, edge computing, machine learning, artificial intelligence, and big data analytics to process patient's clinical healthcare data using blockchain technology. Industrial version 4.0 envisages providing virtualization across personalized healthcare. The concept of blockchain was introduced by Satoshi Nakamoto in 2008. The adoption of blockchain technology can provide security and integrity to EHR thereby improving trust among patients.

## 3 Literature Survey

According to Wright and Bates (2017) Electronic Health Records helps in improving safety, quality and reduces cost in storing patient's data. For maintaining security and privacy, techniques such as cryptography, anonymization and blockchain are adopted in literature. In cryptography, techniques such as hash functions, digital signatures, symmetric and asymmetric keys, and public key certificates are adopted [1]. RFID tags in medical terms provide quick and smart identification of medical records over IoT [2–4]. The sensors enabled with IoT are used to detect anomalies in day-to-day activities [5–7]. Talukder et al. (2018) have proposed an ethereum-based consensus

protocol named proof of disease (PoD) which can solve the challenges raised by EHR [8].

Some of the blockchain algorithms are proof of work, proof of stake, delegated proof of stake, proof of elapsed time, deposit-based consensus, proof of importance, Byzantine fault tolerance, federated Byzantine agreement, hybrid proof of work and proof of stake, and proof of DDoS [9]. Vora et al. (2018) have proposed blockchain approach for secured access and storage of EHRs. Blockchain consists of service contracts, owner contracts, classification contracts, and permission contracts. The authors have proved that by using the cipher manager and adopting encryption/decryption, the unauthorized access can be reduced [10–12].

Yang et al. (2019) have proposed an IoT-based healthcare system with self-adaptive access control. The medical files can be encrypted to provide security, whereas decryption is a time-consuming process in cases of emergency. The authors have implemented a twofold access control technique, a cross-domain technique adaptive for normal as well as emergency. Various attribute-based encryption methods are adopted to impose access control over encrypted data [13–22].

According to Swan (2012), one of the biggest advantages in the usage of IoT is the availability of a large number of low-cost sensors with increased functionalities [23–28]. In sensor networks, public operations are performed using the RSA cryptosystem and design protocols [29].

Blockchain technology improves the traceability of drugs for their originality and helps in investigating foodborne outbreaks. The solution for data privacy and security could be done by blockchain technology. Blockchain technology protects data from failure and data privacy. The blockchain can store all transactional data and the blocks within the blockchain relate to each other in the form of a chain. Each block is identified with a unique hash address which is present in the header. This hash is unique and is generated by the secure hash algorithm (SHA-256) algorithm. The algorithm accepts any plain text and calculates a key of size 256. Each header of the block contains the address of the previous block in the chain. The blockchain has the advantage of not allowing any information to get deleted from the blocks. The major challenge of deploying blockchain in MoT requires high computational power, low scalability, and long latency. Dwivedi et al. (2019) have adopted double encryption using algorithms such as ARX ciphers and public encryption schemes [13]. For the secure transmission of cryptographic keys across a public network, the authors have adopted a Diffie–Hellman key exchange technique. Thus, the above technique is considered a lightweight adaptation technique. Frames and tools such as Hyperledger Fabric, Composer, Docker Container, Hyperledger Caliper, and the Wireshark capture engine are used to improve the performance of Blockchain technology [30].

## *3.1 Blockchain Technology*

The blockchain is a simple digital platform for recording and verifying transactions so that blockchain technology is trending in recent times, since it stores and shares data

in a distributed trust, and immutable manner. Blockchain removes the intermediaries and facilitates a less challenging method for providing access to ledger-based transactions connected with a network. It connects multiple nodes with varying computing powers thereby improving scalability. The techniques and services are supported by blockchain technology.

(a) **Consensus protocol**: Blockchain technology uses healthcare experts, pharmaceutical companies, and insurance agencies to provide a secure environment that can provide data transparency and accountability. Ethereum provides structured data in a secured manner among healthcare sectors with the help of decentralized databases. This structure is designed in such a way that it can protect patient data with integrity and confidentiality. Ethereum provides a platform for the patients through which they have complete control over granting access to particular data to a specialized doctor. The consensus protocol adopted is the ethereum based on future ready proof of disease (PoD) consensus protocol which can identify the disease accurately.
(b) **Provenance**: This is a shared or private immutable ledger used for recording the history of data from its inception throughout its life span. The provenance adopted for this study is the IoT data provenance framework. Data provenance can ensure trust among IoT infrastructure.
(c) **Immutable ledger**: Once a block is written into a blockchain, it cannot be changed.
(d) **Distributed P2P network**: Peer-to-peer network known as a P2P network is decentralized network communication. Nodes are connected at random with the components in the pipeline.

## 4 System Architecture

The blockchain-based patient care monitoring system is explained with four participants such as system admin, patient, doctor, and report. In this system, various functions such as CreatePatientRecord, GrantAccessToDoctor, GrantAccessToLab, RevokeAccess are shown in Fig. 2.

Patients can register through the client application and get enrolled in a certificate authority via a membership service provider (MSP). Upon enrollment, the certificate authority issues the certificate to the patient along with a private key. All transactions to be processed are carried through the Hyperledger blockchain network. Doctors can access those records that are granted permission by the patients. The system admin creates a patient with the required key attributes. The patient in turn can edit the record and complete his profile using the client application, which in turn invokes the chain code for committing a transaction. After committing the transaction, the updated transactions are distributed over the network, so that the data remains in an immutable manner. Transactions are added by a hash value, connected with the previous block.

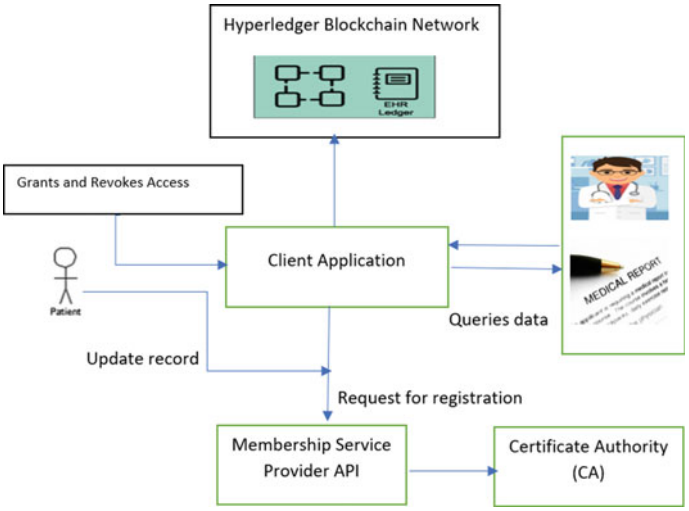



**Fig. 2** System functionality

## *4.1 A Scenario in the Patient Health Monitoring System*

Consider a scenario of a hospital that is monitoring their patients. Sensors like temperature sensors, heartbeat sensors, peripheral pulse oximetry, and blood pressure sensors are attached with a Raspberry Pi microchip. The readings are stored in a database. The previous history of the patients, their alignments, allergic to medicines, previous doctor seen, diagnosis done are in records. It will be better if the current situation of the patient can collaborate with his previous history. These requirements can be incorporated well with blockchain technology with can improve the efficiency and coherence of healthcare delivery. Fundamental issues in the healthcare industry which include lack of data management and distribution can be overcome by blockchain technology which can provide automated, immutable, aggregated, and secure data.

## *4.2 Proposed Algorithms*

The patient care monitoring system deals with four participants, namely system admin, patient, doctors, and reports. The procedure adopted by the system admin is depicted as an algorithm in Table 1. The system admin checks whether the participant (patient or doctor) is valid and requests with the certificate authority to register the participant and issues a public key, if the participant is valid. The admin has complete control over the system to add, read, update, and delete participants. If the activities of the participant are invalid, the admin can block him over the Hyperledger blockchain network.

**Table 1** Algorithm for system admin

```
Input:   Credentials given by the participants (Id, Name, Phoneno, Address)
         Registration certificate requested from Certification Authority (CA)

Output:  Granting access to Patient (P), Doctor (D), Report generator (R)
Procedure SystemAdmin(CA)
  while (True) do
      if (P is valid) then
          add Patient (P_ID, P_name, P_phoneno, P_address) to blockchain network
          grantaccess (P_ID, P_pvtkey)
       else
          invalid participant (P_ID)
       endif
      if (P_ID) is valid
          add Doctor (D_ID, D_name, D_specialization) to blockchain network
          grantaccess(D_ID, D_pvtkey)
      else
          invalid participant (D_ID)
      endif
      if (R_ID) is valid
          add Doctor (R_ID, R_type, R_summary) to blockchain network
          grantaccess(R_ID)
          else
            invalid report (R_ID)
          endif
 end while
 for (all transactions) do
    if found (illegal)
        block (P_ID, D_ID, R_ID)
    endif
end for
end Procedure
```

The procedure to illustrate the role of a patient is depicted in Table 2. The patient requests a private key from the system admin. After being granted access, the patient has access to doctors as well as reports. If PID is in the Hyperledger, the concern person can grant access to the doctor for reading and writing the reports.

The procedure adopted in the working of the module doctor is depicted in Table 3. The input deals with the request given by the doctor to the system admin to enable login. The output is the accessing grant given by the system admin to access the

**Table 2** Algorithm for patient access

Input : Patient ($P_{ID}$) and Key request from system admin
Output: Granting access to $P_{ID}$ transactions in Blockchain network (BN)

```
Procedure Patient (PID)
        while (true)
          if (PID € BN) and  grantaccess(PID € DID)
                grantaccess_read_update (PID, DID, RID,BN)
          else if (PID € BN)
                grantaccess_read_write (PID, DID, RID,BN)
          else
                invalid (PID)
          endif
end Procedure
```

**Table 3** Algorithm for doctor access

Input : Doctor ($D_{ID}$) and Key request from system admin
Output: Granting access to Patient ID ($P_{ID}$) and Report ID ($R_{ID}$) in Blockchain network (BN)

```
Procedure Doctor (DID)
        while (true)
          if (DID € BN) and  grantaccess(PID € DID)
                grantaccess_read_update (PID, DID, RID,BN)
          else if (DID € BN)
                grantaccess_read_write (PID, RID, BN)
          else
                invalid (DID)
          endif
end Procedure
```

patient as well as the doctor. Based on the permission provided by the patient, the doctor will be able to update/access the reports.

## 5 Results and Discussion

For implementation, data recorded from 20 patients was considered. The patient's vital symptoms like temperature, blood pressure, and oxygen saturation level were


**Fig. 3** Screenshot representing creation of blocks in blockchain

monitored and this information is kept on the patient's table. The doctor table contains details of doctors with their specialization and the report table contains the lab reports and ECG reports of the patients.

It is possible to create the functions of the blockchain using flask microframework. The scripts generated with flask can be decentralized by making them run on multiple machines. The data is stored as blocks and the blocks are linked with each other using cryptographic hashes. In this application, a simple interface for sharing data through the posting is created. A post consists of the author's name, content, and time stamp. Every transaction is considered as a block. Each block will have a unique ID. The unique ID is created by the SHA 256 hash function. This hash id is then used to identify the block. The blocks containing transactions are generated one by one and added to the blockchain. If both input and hash are known, the input can be passed through the hash function to proceed to the next block.

In our example, one instance of the blockchain is made to run at port 8000. Two more nodes are located at port 8001 and 8002, so that the new nodes are unaware of the base node 8000 and able to participate in the mining process. When all the nodes are created and transactions are complete, the nodes in the network have complete a chain. A conversation between the patient → doctor, doctor → patient, patient → report is created by creating three blocks that are linked with the hash functions as depicted in Figs. 3 and 4.

## *5.1 Advantages*

**The advantages of our existing model are detailed below:**

A **Automated**: The whole architecture is completely automated requiring almost no human intervention to function except during installation. The sensors collaborate and analyze the data which is then passed over to databases with alerts and reports generated as and when required. An automated system is a lifesaver

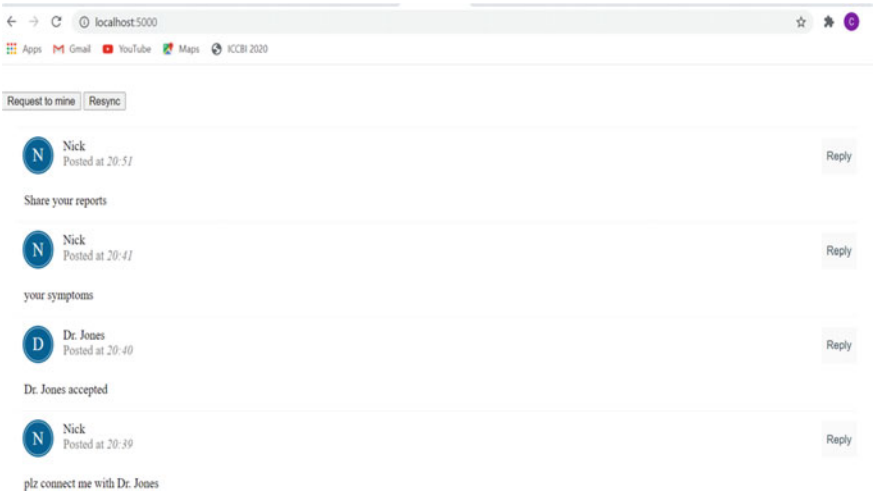



**Fig. 4** Screenshot representing blocks generated in blockchain

for the patient and a relief for the healthcare agencies from which both parties can mutually co-benefit.

B **Robust and versatile**: The whole architecture has been built to be robust and versatile as it can handle the data inflow with a great degree of ease and can be easily implemented in real-time. In the future, several sensors and data nodes can be attached to the microcontroller chip to generate several parameters that can be looked into for data generalization.

C **Cheap**: The whole architecture is cheap in setup and can be implemented seamlessly without any difficulty. This allows even the local healthcare agencies in small towns and villages to set up this system without going into much of the complexities and allow better care for the patients.

D **Real-time monitoring**: Our proposed system allows the healthcare agencies to perform their operations in real-time with the data collected by the sensors, which is a step over the vintage frameworks which required heavy take care and high maintenance costs. With real-time monitoring, the error percentage has also been eliminated uniquely.

## 6 Conclusion

In the proposed system, the model for setting up an Internet of things framework consisting of a Raspberry Pi microcontroller chip and sensors which can collect data from the patient and automate the process by displaying results and inferences from the data collected. This system is expected to revolutionize healthcare management by relaying real-time data to a Web Server/Mobile Application which can be reviewed using a login system for healthcare staff. Further to diagnose the patient efficiently,

patients, doctors, and reports are connected with blockchain technology. Every block is linked to the previous block with the hash field. The consensus proof of disease is adopted for diagnosing the patient. Since the order of transactions is linked with the chain, the data once stored will not be tampered. This application in the future can be implemented with live data and the scalability can be checked.

## References


1. Wright A, Bates DW (2017) Outpatient clinical information systems. Adv Clin Inf 31–50
2. Santos A, Macedo J, Costa A, Nicolau MJ (2014) Internet of things and smart objects for M-health monitoring and control. Procedia Technol 16:1351–1360
3. Wan J, Al-awlaqi MAAH, Li MS, O'Grady M, Gu X, Wang J, Cao N (2018) Wearable IOT-enabled real time health monitoring system. EURASIP J Wireless Commun Networking 298
4. Gomez J, Oviedo B, Zhumo E (2016) Patient monitoring system based on internet of things. Procedia Comput Sci 83:90–97
5. Jara AJ (2014) Wearable internet: powering personal devices with the internet of things capabilities. In: IEEE international conference on identification, information and knowledge, in the internet of things. https://doi.org/10.1109/iiki.2014/9
6. Aminian M, Naji HR (2013) A hospital healthcare monitoring system using wireless sensor networks. J Health Med Inform 4:121. https://doi.org/10.4172/2157-7420.1000121
7. Aldeen YAAS, Salleh M, Aljeroudi Y (2016) An innovative privacy preserving technique for incremental datasets on cloud computing. J Biomed Inform 62:107–116
8. Talukder AK, Chaitanya M, Arnold D, Sakurai K (2018) Proof of disease: a blockchain consensus protocol for accurate medical decisions and reducing the disease burden. In: IEEE SmartWorld, ubiquitous intelligence and computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people and smart city innovations
9. Kombe C, Dida MA, Sam A (2019) A review on healthcare information systems and consensus protocols in blockchain technology. Int J Adv Technol Eng Explor 5(49):473–483
10. Vora J, Nayyar A, Tanwar S, Tyagi S, Kumar N, Obaidat MS, Rodrigues JJPC (2018) BHEEM: a blockchain-based framework for securing electronic health records. 978-1-5386-4920-6/18
11. Swan M (2012) Sensors mania! the internet of things, wearable computing, objective metrics and the quantified self 2.0. J Sens Actuator Networks J(3):217–253
12. Gia TN, Ali M, Dhaou IB, Rahmani AM, Westerlund T, Liljeberg P, Tenhunen H (2017) IoT-based continuous glucose monitoring system: a feasibility study. Procedia Comput Sci 109C:327–334
13. Dwivedi AD, Srivastava G, Dhar S, Singh R (2019) A decentralized privacy-preserving healthcare blockchain for IOT. Sensors. https://doi.org/10.3390/s19020326
14. Yoon EJ, Kim C (2013) Advanced biometric-based user authentication scheme for wireless sensor networks. Sens Lett 11(9):1836–1843 (257–262)
15. Das K (2016) A secure and robust temporal credential-based three-factor user authentication scheme for wireless sensor networks. Peer-to-Peer Netw Appl 9(1):223–244
16. Wong KHM, Zheng Y, Cao J, Wang S (2016) A dynamic user authentication scheme for wireless sensor networks. In: Proceedings of the IEEE international conference on sensor networks, ubiquitous, trustworthy campus (SUTC), Taichung, Taiwan, pp 1–6
17. Das ML (2009) Two-factor user authentication in wireless sensor networks. IEEE Trans Wireless Commun 8(3):1086–1090
18. Farash MS, Chaudhry SA, Heydari M, Sadough SMS, Kumari S, Khan MK (2017) A lightweight anonymous authentication scheme for consumer roaming in ubiquitous networks with provable security. Int J Commun Syst 30(4):e3019

19. Watro R, Kong D, Cuti S, Gardiner C, Lynn C, Kruus P (2004) TinyPK: securing sensor networks with public key technology. In: Proceedings of the 2nd ACM workshop security Ad Hoc sensor networks, Washington, DC, USA, pp 59–64
20. Odelu V, Das AK, Goswami A (2015) A secure biometrics-based multiserver authentication protocol using smart cards. IEEE Trans Inf Forensics Secur 10(9):1953–1966
21. Huang X, Chen X, Li J, Xiang Y, Xu L (2015) Further observations on smart-card-based password-authenticated key agreement in distributed systems. IEEE Trans Parallel Distrib Syst 25(7):1767–1775
22. Choi Y, Lee Y, Won D (2016) Security improvement on biometric based authentication scheme for wireless sensor networks using fuzzy extraction. Int J Distrib Sens Netw 12(1). 8572410
23. Swan M (2012) Sensor mania! the internet of things, wearable computing, objective metrics and the quantified self 2.0. J Sens Actuator Networks 217–253
24. Pereira ORE, Caldeira JMLP, Shu L, Rodrigues JJPC (2014) An efficient and low-cost windows mobile BSN monitoring system based on TinyOS. Telecommun Syst 55(1):115–124
25. Ziegler S, Nikoletsea S, Krco S, Rolim J, Fernandes J (2015) Internet of Things and crowd sourcing—a paradigm change for the research on the Internet of Things. IEEE second forum on internet of things. https://doi.org/10.1109/wf-iot.2015.7389087
26. Routray SK, Anand S (2017) Narrowband IoT for healthcare. In: IEEE international conference on information communication and embedded systems. https://doi.org/10.1109/icices.2017.8070747
27. Shaikh Y, Parvati VK, Biradar SR (2018) Survey of smart healthcare systems using internet of things (IoT). In: IEEE conference on communication, computing and internet of things. https://doi.org/10.11091/ic3iot.2018.8668128
28. Guan Z, Lv Z, Du X, Wu L, Guizani M (2019) Achieving data utility-privacy tradeoff in internet of medical things: a machine learning approach. Future Gener Comput Syst 98:60–68
29. Gupta R, Gupta KK (2017) Patient health monitoring system based on Internet of Things. In: Fourth international conference on image information processing
30. Tanwar S, Parekha K, Evans R (2020) Blockchain-based electronic healthcare record system for healthcare 4.0 applications. J Inf Secur Appl 50:102407

# IoT Based Smart Transport Management and Vehicle-to-Vehicle Communication System


**Vartika Agarwal, Sachin Sharma, and Piyush Agarwal**

**Abstract** Vehicle-to-vehicle (V2V) communication is an advance application and thrust area of research. In the current research, the authors highlighted the technologies which are used in V2V communication systems. Advantage of such technology is that it helps to detect live location and tolling. It plays an important role if there are huge amount of traffic. The current research work can obtain more information about Li-Fi, RFID, VANET, and LORAWAN technology. Li-Fi is known as VLC communication system that uses visible light for high data transmission and reception. RFID technology helps the emergency vehicle to reach destination quickly by avoiding any kind of traffic. LORAWAN is a large-scale network technology with a long range and VANET with low power that allows to obtain accurate traffic information on each route and this saves time. The comparison between the different technologies is reviewed in order to obtain the optimized technology as per the applications.

**Keywords** Internet of things (IoT) · Vehicle-to-vehicle (V2V) communication · Li-Fi · RFID tag

## 1 Introduction

IoT is one of the most significant, dynamic, and groundbreaking research contributions of the twenty-first century [1]. IoT has brought a drastic change in the lifestyle of every individual through its vast range of applications. The best part in the current development is the storage of data in cloud. The use of IoT based application is implemented in almost all the fields. Transport management system has several limitations and is constrained by existing technologies. Advance tools like Li-Fi, RFID, LORAWAN, and VANET are discussed in the current research. Currently, the use

V. Agarwal (✉) · S. Sharma · P. Agarwal
Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India

S. Sharma
e-mail: sachin.cse@geu.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_55

of IoT with Li-Fi technology is implemented and executed in transport management system by using V2V communication system. Authors have described the effectively the use of Li-Fi in V2V communication [2]. This technology uses several input and output devices. The use of Li-Fi has created the easiness in the data transmission and storage capacity. The hardware and software components, Li-Fi communicator system, etc., are required for Li-Fi technology which are discussed below.

### *1.1 Hardware Component Needed for Li-Fi Technology*

Following are the conventional hardware components used in the execution of Li-Fi Technology:

- **Sensor**—Sensor is useful to find out the speed and distance of vehicle.
- **LED Light (Transmitter)**—White LED will be used for data transmission from vehicle 1 to vehicle 2.
- **Photodiode**—Photodiode will be used for data detection.
- **Wi-Fi Module**—The proposed system requires efficient communication using the internet.

### *1.2 Software Used for Li-Fi Technologies*

Following are some of the software which are very useful in the development and execution of Li-Fi Technology:

- **Arduino IDE**—Arduino is used to install Arduino UNO board. Program written using Arduino software is called sketches.
- **Telegram App**—Telegram App is used to directly send the accident location to the hospitals and police station.

### *1.3 Li-Fi Communicator Systems*

Li-Fi communication system as shown in Fig. 1 which consists of accelerometer, Arduino UNO, cloud intimating device, ultrasonic sensors, Li-Fi transmitter and receiver, LCD display, node MCU, DC motor, etc. Some important points that need to be considered while using Li-Fi for transport management system includes:

(a) Activation of sensors and actuators is necessary to enable vehicle-to-vehicle communication between two vehicles.
(b) The sensor reads the received data, interprets it, and then transmits it in the suitable form to another vehicle under communication.
(c) The complete data is processed by microcontroller.

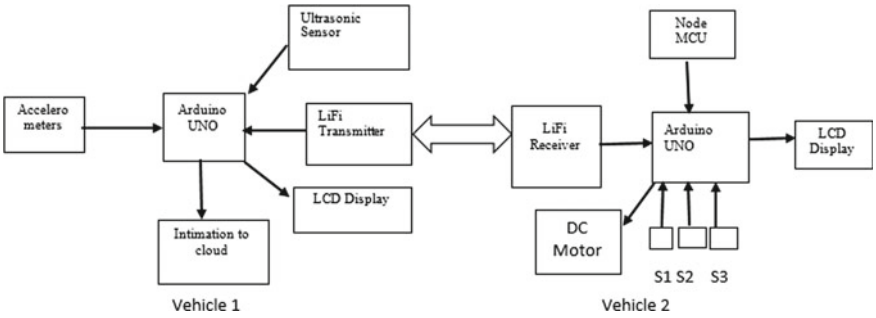



**Fig. 1** Systematic block diagram for Li-Fi communication

(d) The updated and real-time modification is done by using cloud-based technology.

### *1.4 Limitation of Li-Fi Technology*

When it comes to technological advancements, there are several limitations associated with the technology. Following are some of the limitations of Li-Fi technology which can be treated as further area of research in the field of V2V communication.

(a) Li-Fi works on limited light range (suitable under visible light only).
(b) Interference in Li-Fi occurs in sunlight.
(c) The major drawback in Li-Fi fidelity is that it shows poor performance in dark.

For fulfilling the limitation of Li-Fi, radio-frequency identification (RFID) tags are used for leveraging vehicle-to-vehicle communication. It is also considered as an important cloud-based tool.

## 2 Radio-Frequency Identification (RFID) Tag Technology

RFID tag is used for transmitting and receiving information via an antenna and microchip. It is also called as integrated circuit (IC). RFID tag is used for live tracking, tolling, and real-time vehicle tracking application. Many researchers have effectively highlighted the design and working of RFID tag to track data in V2V communication application [3]. They are small in size, lightweight, and can potentially last a lifetime.

### 2.1 Working of RFID Tag

According to this system, all vehicles have radio frequency identification tags placed on them. RFID reader reads tag from vehicle and turns green. It is basically used to count no of vehicles passed through signal [4]. When emergency vehicle passes on the road, it will convert into red.

### 2.2 Limitations of RFID Tags

There are several limitations of RFID tags as compared to other labels. Though it has wide range of applications, RFID tags needs lots more improvement in terms of security. Some of the limitations are mentioned below:

- RFID tags are not suitable for no reasons including protection and technical problems compared to other marks.
- RFID tags are not possible for readers to differentiate. Almost everyone can read details.

### 2.3 Vehicle to Everything Communication (V2X Communication)

IoT based vehicle to everything communication is another good technology which is very helpful in vehicle navigation, vehicle transportation management, and vehicle security management. The insights for this are mentioned below [5].

(a) **Communication Type Insights**—V2I and V2V play an important role if there are huge amount of traffic.
(b) **Connectivity Type Insights**—Such systems are being installed in cars.
(c) **Vehicle Type Insights**—Scope of vehicle to everything technology is increasing day by day.

For connecting to the cloud, LORAWAN is used. With the help of LORAWAN, the parking area in which the vehicle has to be parked can be easily identified.

## 3 LORAWAN Technology

LORAWAN, a long range, low-power wide area networking technology, enables several advance services, which are not possible through other network-based technologies [6]. LORAWAN provides best battery life, and it is very economical for a wide area network due to its reduced cost and good battery life.

### *3.1 Implementation of Proposed Framework*

LORAWAN implementation requires a specific set of framework for implementation [7].

**V2V Communication**—Beacon is used for scanning nearby devices. In this, the first 8 bit consists of manufacturer code, the next 20 bits has the vehicle information, the next 2 bits has the information about type of vehicle, and the last 2 bits shows the emergency situation.

**UUID**—Universally Unique Identifier (UUID). Further, it can be divided into five groups.

### *3.2 Proposed Framework*

Lights are used for scanning surrounding vehicles. For instance, if one vehicle journeys steadily and the second vehicle travels at high speed. With the help of a beacon, an alert message will be passed to second vehicle. Hence, collision is avoided, and the vehicle can pass easily without any interference.

**V2I communication**—With the help of V2I communication, information can be collected from the cloud.

**V2V communication**—Here, transmission is possible between nearby vehicles. It had some major problems like Bluetooth low energy consume less power.

**Limitations of LORAWAN**

LORAWAN has several limitations, where it can only send small packets through LORAWAN. For large packets, it is not suitable. LORAWAN is not suitable if you want to control light in your house.

## 4 Vehicular Adhoc Network (VANET) Technology

VANET is a developing technology nowadays [8]. It helps an emergency vehicle to select best routes with no traffic for reaching the destination easily. System also generates a warning message if an emergency vehicle meets an accident.

### *4.1 Working of VANET*

VANET can be used in simple cases as shown in Figs. 2 and 3. In Fig. 2, an emergencies vehicle is initially at rest in the parking lot of the hospital. Following this, the vehicle allocated a destination of urgency with two separate roads. Emergency vehicles should select the best route in order to reach their destination fast [9].

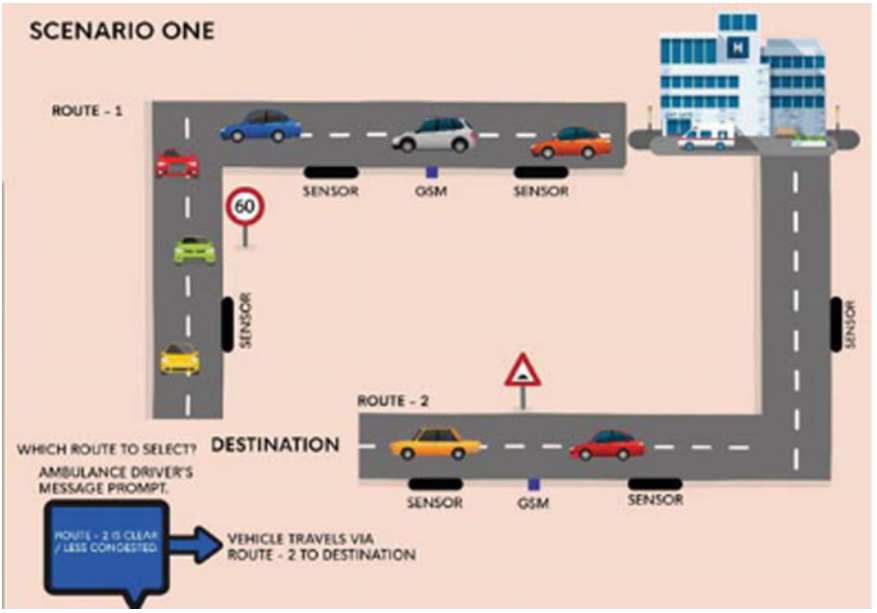



**Fig. 2** Case 1: traffic management system testing [8]

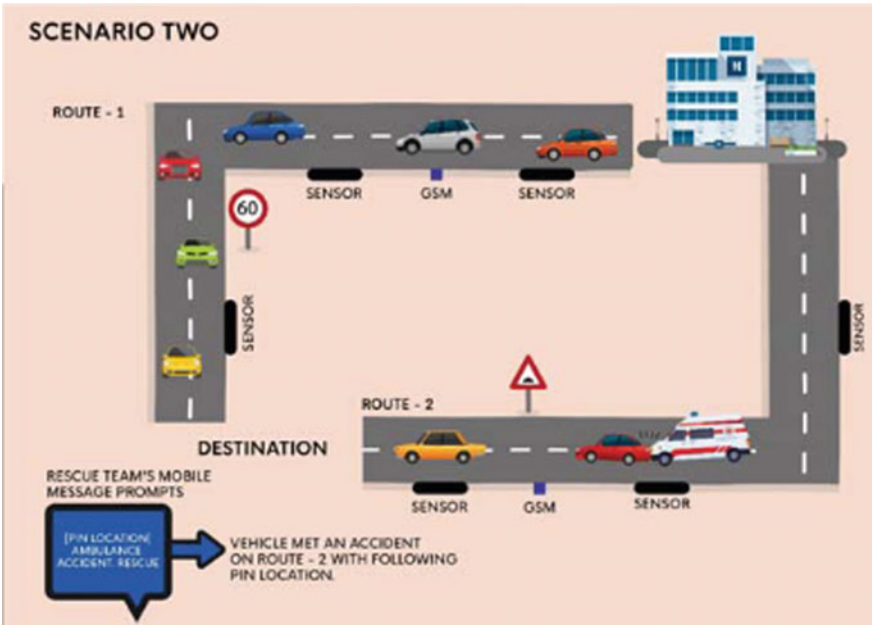



**Fig. 3** Case 2: traffic management system testing [8]

Driver can know more about best route by sending a message 'route status.' The optimized and less traffic route can be identified. When the message is sent, GSM will send information to the hospital informing about best route which is more preferable. Vehicle now travels in the best route. The use of this technology enables the driver to provide the fastest, safe, and least traffic route which saves life, money, and time of the complete process [10].

Figure 3 highlights the scenario 2 for traffic management system testing where different sensors and GSM module are installed in the vehicle under communication to rescue the uneven happening through mobile messaging. This is even advanced concept then simple IoT. In case, if an emergency vehicle meets an accident, a message with location will be automatically sent to the relevant rescue authorities [11].

### *4.2 Emergency Application for V2V Communication*

In an emergency application for vehicular network, warning messages included auto collisions, street traffic, and so on, and then, large number of duplicate packets congest networks [12, 13]. The V2V communication works under ultra-low latency of data and information sharing among users [14, 15]. Several corporate and business issues can be easily rectified using the advance industry 4.0 based technology. Reduction in response timings can be systematically analyzed in V2V communications.

### *4.3 Vehicle to Infrastructure Communication*

It is a combination of V2V and V2X, and vehicle to infrastructure communication is the communication from accident vehicle to base station. The message is for emergency situation which forwards it to the nearby rescue center. The architecture also works in security management along with the traffic management system.

## 5 Conclusion

The current manuscript reviewed on the applications of IoT in vehicle-to-vehicle communication system. Different technologies like Li-Fi Technology, RFID tag, VANET, and LORAWAN technology are highlighted along with its limitations and applications in vehicle-to-vehicle communication system. Also the limitations of all the systems are highlighted in the research. The amalgamation of different IoT based tools can finally summarize the utility based on the application, specification, and the nature of use. The further scope of research is being tried to identify from the

limitations in existing technology and infrastructure. Also the use of V2V communication using advance Industry 4.0 tools can reduce the efforts, time requirements, and money in different monitoring and security applications. Data capturing and modification is optimized using these technologies.

## References

1. Chavhan S, Gupta D, Chandana BN, Khanna A, Rodrigues JJ (2019) IoT-based context-aware intelligent public transport system in a metropolitan area. IEEE Internet Things J
2. George R, Vaidyanathan S, Rajput AS, Deepa K (2019) LiFi for vehicle to vehicle communication–a review. Procedia Comput Sci 165:25–31
3. Uddin MJ et al (2009) Design and application of radio frequency identification systems. Eur J Sci Res 33(3):438–453
4. Wyld DC (2006) RFID 101: the next big thing for management. Manage Res News
5. Ullah H et al (2019) 5G communication: an overview of vehicle-to-everything, drones, and healthcare use-cases. IEEE Access 7:37251–37268
6. Adelantado F, Vilajosana X, Tuset-Peiro P, Martinez B, Melia-Segui J, Watteyne T (2017) Understanding the limits of LoRaWAN. IEEE Commun Mag 55(9):34–40
7. Reynders B, Wang Q, Pollin S (2018) A LoRaWAN module for ns-3: implementation and evaluation. In: Proceedings of the 10th workshop on ns-3, pp 61–68
8. Syed MSB et al (2020) IoT based emergency vehicle communication system. In: 2020 International conference on information science and communication technology (ICISCT). IEEE, pp 1–5
9. Sharma S et al (2018) A security system using deep learning approach for internet of vehicles (IoV). In:2018 9th IEEE annual ubiquitous computing, electronics and mobile communication conference (UEMCON). IEEE, pp 1–5
10. Gupta N et al (2020) SDNFV 5G-IoT: a framework for the next generation 5G enabled IoT. In: 2020 International conference on advances in computing, communication and materials (ICACCM). IEEE, pp 289–294
11. Sharma S et al (2019) Blockchain-based internet of vehicles (IoV): an efficient secure Ad Hoc vehicular networking architecture. In: 2019 IEEE 2nd 5G world forum (5GWF). IEEE, pp 452–457
12. Sharma S, Mohan S (2020) Cloud-based secured VANET with advanced resource management and IoV applications. In: Connected vehicles in the internet of things. Springer, Cham, pp 309–325
13. Sharma S et al (2019) Advanced spectrum management for next-generation vehicular communication: an AI approach. In: 2019 IEEE 10th annual information technology, electronics and mobile communication conference (IEMCON). IEEE, pp 0632–0637
14. Sharma S et al (2018) Smart vehicular hybrid network systems and applications of same. U.S. Patent application 15/705,542, filed March 29
15. Sharma S, Mohan S (2016) Dynamic spectrum leasing methodology (DSLM): a game theoretic approach. In: 2016 IEEE 37th Sarnoff symposium. IEEE, pp 43–46

# An Analytical and Comparative Study of Hospital Re-admissions in Digital Health Care


**Aksa Urooj, Md Tabrez Nafis, and Mobin Ahmad**

**Abstract** Medical re-admissions are expensive and indicate poor quality of hospitals. A remarkable re-admission rate has a huge financial impact on the patients and the hospital. Through the increasing use of medical health records, enormous data is available to us for review which can identify high-risk patients effectively and decrease the mortality rate. According to the 2010 Affordable Care Act, hospitals may be penalized for re-admitting patients within 30 days of discharge. However, hospitals claim that the root cause of re-admissions lies similar to the populations served. In this paper, various re-admission prediction techniques that have been used earlier to estimate the re-admission rate of patients after being discharged from the hospital have been discussed. This article will also give a summary of various socioeconomic and demographic factors that play a vital role in medical re-admissions. Moreover, it will cover some of the measures that can be taken to reduce hospital re-admissions.

**Keywords** Digital health · Electronic medical records · e-Health · Hospital re-admission · Intervention · Big data analysis

## 1 Introduction

Hospital re-admission is the unnecessary return of the patient to the hospital following an original admission and discharge from the hospital. For research purposes, various time windows like 30 days, 90 days, and 1 year re-admissions were used. The time period often used is 30 days [1]. Re-admission may take place at the same hospital as the original or to another hospital for planned or unplanned medical treatments [2]. High re-admission rates indicate relatively low performance and have adverse

A. Urooj · M. T. Nafis (✉)
Department of Computer Science and Engineering, Jamia Hamdard, New Delhi, India

M. Ahmad
Faculty of Science, Jazan University, Jazan, Saudi Arabia
e-mail: msyed@jazanu.edu.sa



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_56

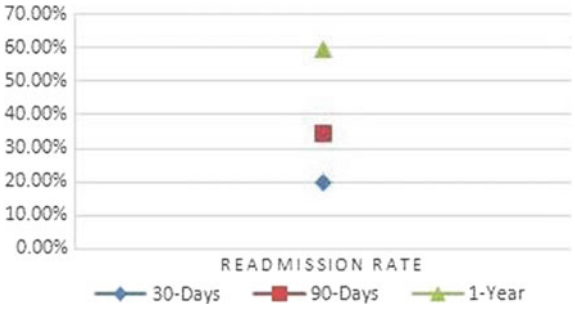



**Fig. 1** Hospital re-admissions within 30 days, 90 days, and 1 year

effects on both patients and hospitals [2]. It is recorded that about 19.6%, 34%, and 56.1% of the patients released from the hospital were re-hospitalized within 30-days, 90-days, and 1-year, respectively, according to a study in the New England Journal of Medicine [3] (Fig. 1).

The study carried out by the Medicare Payment Advisory Commission (MedPAC) showed that within 30 days of discharge from the hospital, approximately 17.6% of patients are readmitted, reflecting $35 billion of medicare expenditure a year [4]. Also, around 75% of the re-admissions are avoidable which can save around $17 billion every year [2, 5] (Fig. 2).

As per the Affordable Care Act, hospital re-admission reduction program (HRRP) was established by the US Center for Medicare and Medicaid Services (CMS), which aims to penalize hospitals with immoderate re-admissions within 30 days. Initially, only three conditions were taken into consideration, i.e., heart failure (HF), acute myocardial infarction (AMI), and pneumonia (PN), and three more conditions will be included starting 2015 [6].

According to the HRRP data (Figs. 3, 4, and 5), it is quite evident that re-admissions from 2013 through 2016 dropped marginally, and rates differ based on the type of diagnosis. During 2015 and 2016, re-admission rates for acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN) remained the same, however [7].

CMS calculated the re-admission rate from 2008 to 2011 discharge data and used it as a benchmark which changes every year by considering the previous three years


**Fig. 2** Overall re-admission expenses per year

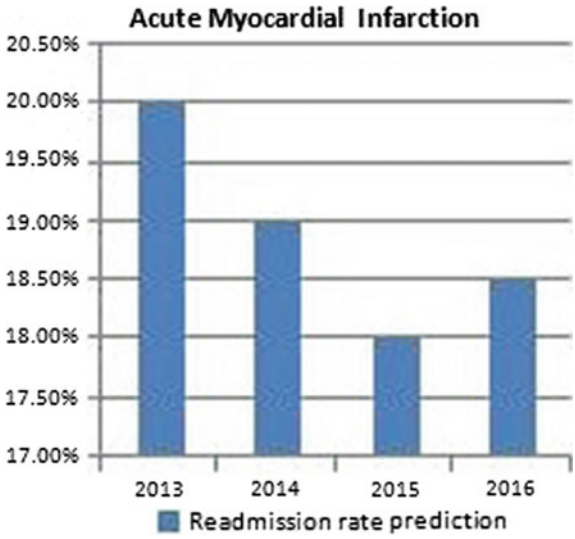



**Fig. 3** Re-admission rate for AMI [7]

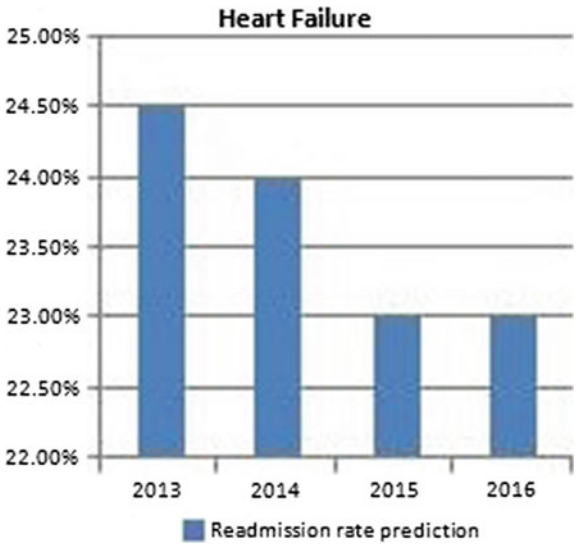



**Fig. 4** Re-admission rate for heart failure [7]

[8]. The fine for each hospital was fixed as 1% of the overall medicare payments from October 1, 2012, and will gradually increase to 3% in 2015 and beyond. Noticeably in 2013, at least 2000 hospitals were reimbursed, with a gross forfeited amount of $280 million [9] (Fig. 6).

This article is organized as follows. In Sect. 2, a brief overview of the techniques previously used for estimating hospital re-admissions is presented. In Sect. 3, a summary of the causes responsible for high re-admission rate is discussed. Section 4

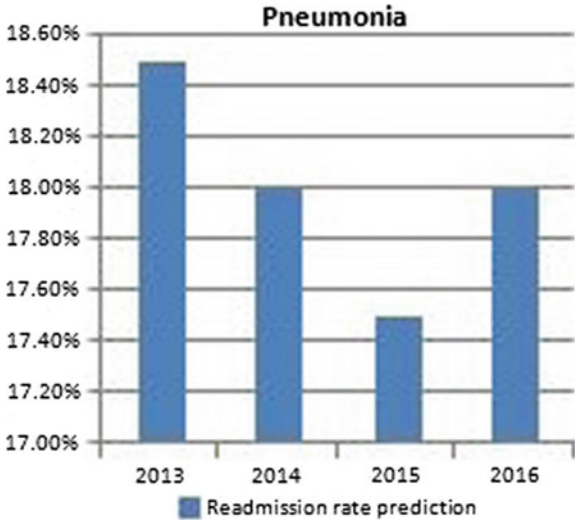



**Fig. 5** Re-admission rate for pneumonia [7]

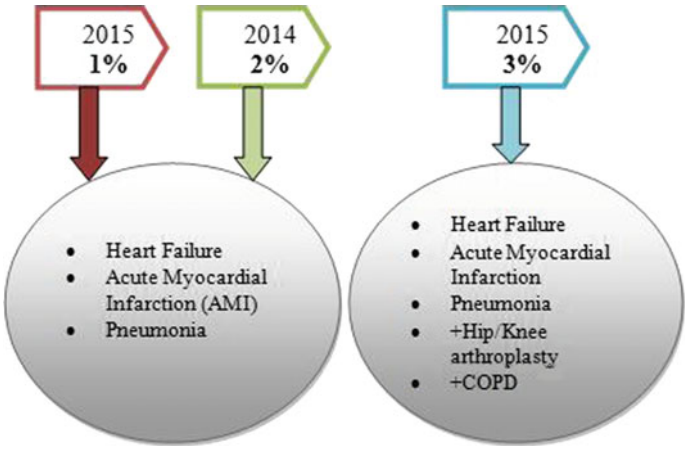



**Fig. 6** Hospital re-admission penalties

introduces some of the ways to reduce hospital re-admissions. Finally, Sect. 5 draws the conclusion of the research work and future scope.

## 2 Existing Work on Re-admission Prediction

Re-admissions after procedures are expensive and indicate the poor quality of a hospital. There are a large number of prediction algorithms that use clinical and administrative data to the forecast re-admission risk of patients [10–12]. To estimate the re-admission risk of a patient, dedicated efforts are embedded in the learning of effective features that are most relevant to a patients' re-admission. A short summary of research work performed earlier to predict re-admissions at hospitals was discussed.

1. Risk-O-Meter is a standalone system used to forecast the re-admission risk factor of CHF patients [13]. However, it is only accessible via a Web interface and lacks the ability to interface with other systems.
2. Re-admission score as a service (RaaS) is the first re-admission risk calculator. It leverages cloud computing to estimate the 30-day re-admission score of heart failure (HF) patients and also displays the leading factors responsible for their re-admission. The main advantage of RaaS implementation in comparison with previous efforts is that it is deployed on Microsoft Azure which provides access to both clinicians and patients without any requirement of analytical infrastructure. Moreover, it is flexible enough to integrate with other medical systems and data sources [5].
3. To reduce penalties and increase the life span of patients, [1] used clustering-based actionable knowledge mining to decrease the average number of re-admissions. First of all, patients are clustered based on the similarities in their diagnoses, and then, novel algorithms are applied to predict the order of treatments that the patients have to undergo to reach the desired condition. Finally, a scoring metric is used to evaluate the clusters and treatments, so that a better procedure can be estimated for the new patient, which decreases the hospital re-admission rate to a large extent.
4. Some traditional modeling approaches, for example, logistic regression or support vector machine (SVM), is commonly used for classification problems [14–17]. Work done by Ref. [18–20] result in more statistical features that are used to enhance re-admission prediction using modern modeling methods. However, most of these methods focus on the ability to extract and classify features, which restricts the efficiency of their methods.
5. A data-driven approach was implemented for forecasting the re-admission of hospitals based on administrative results [21]. However, this technique cannot integrate clinical laboratory data into the model and thus cannot equate its results explicitly with other methods. Thus, [22–24] used a range of data sources covering patient social and demographic profiles, medications, treatments, illnesses, and laboratory tests. They used the features built for a particular illness and failed to consider the negative impact of misclassification errors, however.
6. Choi et al. [25] exhibit deep learning models that outperform conventional modeling methods in the medical field and therefore can be interpreted for the

analysis of health care [26]. These works, however, mainly focused on deep learning; do not pay attention to the problem of skewed data.

7. Therefore, [2] used convolutional neural networks (CNN) to extract important characteristics that are most relevant to hospital re-admission. They used categorical feature embedding to encode clinical features with feature vectors. Then, both these categories of features were fed into the multilayer perceptron (MLP) as input. To overcome the imbalanced data problem, a cost–sensitive deep learning model was proposed to train MLP during re-admission prediction. This approach performs better on real clinical datasets than the existing methods and is being used in a large hospital for treatment and decision making on a real system.
8. Most models concentrated on a particular section of the population, for example, a model for classifying heart failure patients at re-admission risk within 30 days was designed in [27] using a comprehensive dataset consisting of 12 separate cohorts of 1100 to 14,500 patients each.
9. Alluhaidan and others [28] designed a telehealth system for heart failure patients to fill the gap as patients come to their homes to improve self-care, and therefore, reduce re-admissions. Similarly, [29] developed a decision tool for assessing and decreasing the chances of re-admission in patients diagnosed with CHF. However, this tool was designed particularly for an urban teaching hospital in the USA.
10. The ongoing studies on medical re-admissions continue to include clinical data which is available merely in the advanced electronic medical records (EMRs). It is troublesome because most heart disease victims belong to the population whose hospitals are not able to adopt modern EHR systems rapidly. Vedomske et al. [30] used the random forest machine learning methodology to forecast emergency re-admissions within 30 days of discharge for CHF patients. However, this technique had some drawbacks. First of all, data was taken from a particular hospital, and other locations were not taken into consideration. Secondly, death rates within 30 days of discharge were not taken into account. Lastly, a huge portion of data was excluded due to missing re-admission details.
11. A latest comprehensive and in-depth analysis of models for forecasting hospital re-admissions explaining their efficiency and evaluating clinical suitability is described in [31].
12. This article determines the efficacy of care interventions in older adults with chronic illnesses in minimizing short- and long-term hospital re-admission following discharge from the hospital [32].
13. Likewise, a model was created for older patients over 65 years of age [33]. This experiment utilized medicare inpatients from the general US community, with a comparatively limited collection of data. The aim was to forecast complex 30-day treatment changes. Comparatively, the model gave a good performance (0.83 AUC).
14. This study provides a model for evaluating the optimum time frame when defining re-admissions for patients. Initially, by fitting a special case of a coxian phase-type distribution, which is expressed as a mixture of two generalized

Erlang distributions, it captures the re-admission process. And then, the optimal time period is determined by applying the minimum classification error method [34].

15. Within this article, it demonstrates that a broader study of re-admissions, utilizing clustering, gives a richer and more complex interpretation. Several commonly used U.S. datasets available publicly were combined: medicare inpatient provider utilization, CMS impact file, hospital cost report, hospital compare data, data from re-admissions reduction program (HRRP), payment data, and U.S. census data. The primary hospital characteristics that were mainly related to re-admissions were examined and are given as patient satisfaction, the effectiveness of care, caring for poor, structure measures, case mix, hospital funding, and use of hospital beds [7].
16. This article suggests an analytical framework using hospital inpatient administrative data from a nationwide healthcare dataset. The stacking algorithm is integrated with an updated weight boosting algorithm to form a joint ensemble learning model [35].
17. This project aimed to build a model that predicts the risk of re-admission within 30 days. The risk of re-admission calculated by the proposed model was then tested with the LACE index and patient at-risk of hospital re-admission (PARR)—the most popular models for finding out the risk of re-admission [36].
18. Chronic obstructive pulmonary disease has been recently introduced to the list of diseases for which hospitals in the U.S. are penalized by the CMS. While attempts have been made to statistically forecast those most vulnerable to re-admission, rare studies have worked on unstructured clinical records. It has designed a framework for analyzing clinical records and estimating re-admission using natural language processing [37].
19. A study of five statistical models for COPD re-admission prediction including generalized estimating equations, logistic and Bayesian logistic regression, logistic regression tree, and generalized linear mixed model is discussed in [38]. Hasan et al. [39] present a simple model that uses logistic regression to determine patients at high risk of re-admission.
20. Researchers at Deakin University developed a framework to assess the re-admissions of chronic diseases [40]. The framework can predict the re-admission rate in COPD patients within 30 days with an AUC = 0.67 and was an enhancement upon baseline co-morbidity methods which are commonly used for re-admission prediction.
21. The project proposes a causal Bayesian network model to find ways of reducing re-admissions for patients diagnosed with chronic obstructive pulmonary disease (COPD). This model uses a Bayesian network approach and adopts domain knowledge. This study provides early intervention plans which prove really helpful to reduce COPD re-admissions [41].
22. The main objective of this paper is to develop an optimization model that helps to predict the re-admission rate of chronic obstructive pulmonary disease

(COPD) patients within the budget constraint. This model will also assist the hospitals to plan the required budget and reduce the re-admission levels [42].

23. Moreover, healthcare delivery systems such as nutrition analysis, radiation therapy treatment planning, resource allocation, and scheduling (e.g., [43–47]) heavily used linear programming models. But, no such study has been done on COPD re-admission.
24. This study used various machine learning approaches such as logistic regression decision tree, neural networks, and gradient boosting to classify different demographic, clinical, and socioeconomic variables that play a key role in forecasting the loss of revenue due to re-admissions. For this purpose, three medical conditions were mainly used, namely total knee arthroplasty (TKA), chronic obstructive pulmonary disorder (COPD), and total hip arthroplasty (THA). These models are then tested and contrasted based on area under ROC (AUC) and the rate of misclassification. With the help of visual data, this tool not only helps the hospitals to measure financial costs but also to track their service efficiency in a real-time fashion [8].
25. This study proposes a deep learning model aggregating deep forest and wavelet transform to predict the re-admission risk of diabetic patients. The model has been validated with actual health reports and contrasted with other patient prediction strategies. The findings of the experiment indicate that the deep forest is capable of performing better than the existing state-of-the-art methods for diagnosing diabetics [48].
26. Duggal et al.'s recent research makes use of Apache cTAKES to annotate the unstructured electronic health record (EHR) [49]. This study looks primarily at the re-admission rate of patients with diabetes in an Indian hospital for over 30 days. Several machine learning algorithms such as Adaboost, logistic regression, random forest, neural networks, and naïve Bayes were compared. The highest AUC of 0.688 was achieved using random forests.
27. This paper initiates the study of predicting the risk of re-admission for congestive heart failure (CHF) incidents within 30 days using big data solutions [50].
28. This project develops a software tool that not only calculates the re-admission probability of CHF patients but also assists the hospitals to determine the various factors that have a profound impact on the calculated re-admission risk. Depending on the patient's social factors and medical comorbidities, the decision support tool determines a list of post-discharge resources to reduce the probability of re-admission for patients [29].
29. A detailed list of models for forecasting the heart failure patient's re-admissions risk is given in [51].
30. Research by Wasfy et al. aims to use the unstructured details in the electronic health record (EHR) to forecast re-admission in percutaneous coronary intervention patients within 30 days [52]. In this analysis, the primary tool of NLP was the use of regular expressions to derive particular queries from the clinical report. The AUC for this study was 0.69.

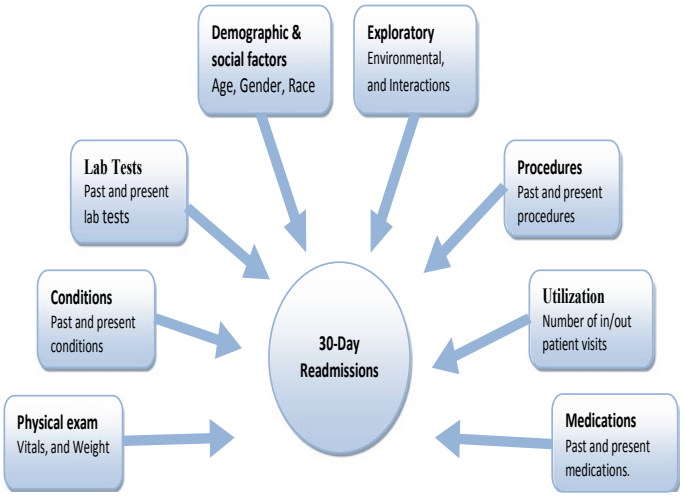



**Fig. 7** Re-admission risk prediction conceptual model

31. Kansagara et al. conducted a systematic review in 2011 which compares data, methodology, and results [53]. The study concluded that estimation of re-admission is a complex issue and that recent models do not outperform research a decade before.
32. In this paper, the authors worked on a general framework for hospital-specific re-admission prediction. It extracts patient data (medications, demographics, labs, CPT and ICD codes, etc.) from a health system and generates a statistical model to predict re-admission. If a model for a particular disease is required, the framework can also adjust the model for only those patients suffering from that disease [6].
33. Some models were also developed for general population environments. A common model named LACE is discussed in [54], which used data from 4000 patients in 11 hospitals to predict emergency re-admissions after discharge from hospitals in Ontario, Canada. Other models that focus on data from general population include [55–57] (Fig. 7).

## 3 Causes of Hospital Re-admissions

Over the past years, hospitals across many developed countries such as the UK and the USA experience high re-admission rates [58]. One reason behind this could be a rapid discharge of the patients, as hospitals are accused of discharging patients "sicker and quicker" [59]. Therefore, the rise in early discharges may lead to high re-admission rates, which could probably be perceived as the premature discharge of patients from hospitals.

However, hospitals believe that re-admissions depend on the type of populations being served [7]. The argument is backed by conventional statistical analyses. Particularly, while the patient population is an important factor, patient satisfaction and quality of care also contribute a lot. Hospitals with higher patient satisfaction rates, serving less poor populations, and dealing with more complicated cases have lower re-admission rates across 1518 hospitals across the United States and are less likely to incur fines, according to research. There are poor associations: The frequency of re-admissions is more for the low-income population, and less for patients with high satisfaction; quality of care is negatively correlated with poor populations; resource availability is positively correlated with complex cases [7]. In short, continuity of care services prevents hospital re-admissions particularly in elderly people with serious illness [32]. On top of that, social characteristics play a vital role in predicting the patient's risk of getting readmitted to a hospital [60].

Drug consumption such as cocaine, alcohol, and tobacco harms re-admission of CHF [61]. It is evident from the fact that CHF patients who use tobacco and/or alcohol are more vulnerable to re-admissions compared to other patients [62]. Several researchers report that patients receiving medication for tobacco dependency have lower chances of re-hospitalization within 30 days of discharge. Similarly, cocaine use is known as a major social element in 30 days of re-admission or death [63].

Recent findings across the general population found smoking to be a leading cause of re-hospitalization. With a re-admission's odd ratio of 1.33 after one year of release, patients with mental disorder plus smoking addiction have a 33% higher probability of re-hospitalization relative to those who never smoked [64].

Social isolation is also found to be responsible for increasing re-admission rates, according to a study [60]. The possibility of re-admissions can be reduced if patients get continued support from their family members [62]. It has been shown that the availability of resources, such as travel arrangements, affects the re-admission rate [65].

Research has shown that there is a lack of responsibility to provide additional medical care for CHF such as patients refusing their diagnosis and declining to continue engagement with their doctors which further affects re-admission rates [65].

Besides, the stress caused by demographic variations, for instance, elderly people in advanced countries, may influence the rise in re-admissions. The main triggers of premature re-admission in the UK are considered to be chronic obstructive pulmonary disorder (COPD), stroke, congestive heart failure (CHF), and hip and thigh fractures [66, 67]. The re-admission rates are high for patients diagnosed with CHF, as per a report released by the American Heart Association (AHA). Such that within 30 days of discharge, 21.2% of overall CHF patients were re-hospitalized [28].

In addition to exploring medical, demographic, and ethnic causes, North Carolina data provided by HCUP found that social and economic factors such as eating habits, sleeping patterns, smoking, diet, and income play a vital role in these re-admissions [8].

Another reason behind patients getting readmitted to hospitals is the undesirable outcome that may occur after the recommended treatments and may not be anticipated

beforehand [68]. Furthermore, statistics demonstrate that 10% of patient fatalities occur due to diagnostic mistakes and are the most common type of litigation for medical malpractice in the USA [69].

According to a study released as a review letter in JAMA, Sepsis is found to a leading cause of medical re-admission and expenses. Nationwide re-admissions database 2013 was used to examine costs of unplanned re-admissions for heart failure (HF), pneumonia, chronic obstructive pulmonary disease (COPD), acute myocardial infarction (AMI), and sepsis in adults in the USA. The predicted re-admission costs for sepsis ($10 070) were significantly higher than for heart failure ($9051), pneumonia ($9533), COPD ($8417), or AMI ($9424) [70].

## 4 Reduction in Hospital Re-admission

The huge expenditure on healthcare each year and the need to improve the quality of healthcare make it mandatory to reduce hospital re-admissions. Early re-admission prediction in the process of hospitalization will help prevent severe or life-threatening accidents, which serve as a significant contributor to decrease medical costs. Previous research on forecasting re-admission mostly concentrates on some important features that are most relevant to the cause of re-admission [71].

A wide research area is working to reduce re-admissions. A large portion of this concentrates on enhancing discharge procedures and quality of care, e.g., ensuring that patients are properly informed regarding their post-discharge care [6]. It is essential to educate patients about their sickness and the various treatment procedures that they can have to maintain a stable state. Research conducted by Sara Paul showed that education can decrease the chances of re-admission by 54% [72].

One study showed that setting up a clinical care program, that includes regular telephonic meetings between the pharmacist and a patient, reduced chances of re-admission by 26% [18], for example, re-admission risk in hospitals for high-risk aged people released from hospital wards can be minimized with home-based pharmacist follow-up that detects and fixes drug-related issues (DRIs), claims Australian researchers.

Specifically, when prolonged re-admissions are related to socioeconomic and clinical determinants, improved coordination and integration through multiple boundaries (health, social, cultural) and structured collaborations between community-based organizations and acute-care hospitals may deter them (Liner-tova et al. 2011) [32].

Existing research has found that improved assistance to patients at home can have a drastic impact on healthcare costs and efficiency [73, 74], for example, [28] designed a telehealth system for patients diagnosed with congestive heart failure (CHF). It collects patient information, for example, glucose, heart rate, weight, blood pressure, etc., and displays it on the dashboard present on the clinician's end. Daily measurement of vital factors (blood pressure, weight, heart rate, glucose) and regular tracking

of patient symptoms can decrease re-admissions and improve patients' condition to a great extent.

Many behavioral considerations, such as non-adherence to drugs and eating patterns, greatly affect the risk of a patient's re-admission [75]. It is advised that an individual with CHF-related problem limit their sodium intake to 3 grams or less [76]. The chances of heart failure that can get doubled due to hypertension can be controlled by reducing the amount of sodium intake [77]. Recently, the new guidelines released by the American Cancer Society suggest that women with acceptable breast cancer risk should begin mammograms at 45 years of age (5 years later than previously recommended by ACS) [78–80]. According to Fogg's behavior model, the three important elements to adopt a healthier lifestyle are motivation, ability, and trigger to change your behavior. If anyone among these is missing, behavior does not change (Fig. 8).

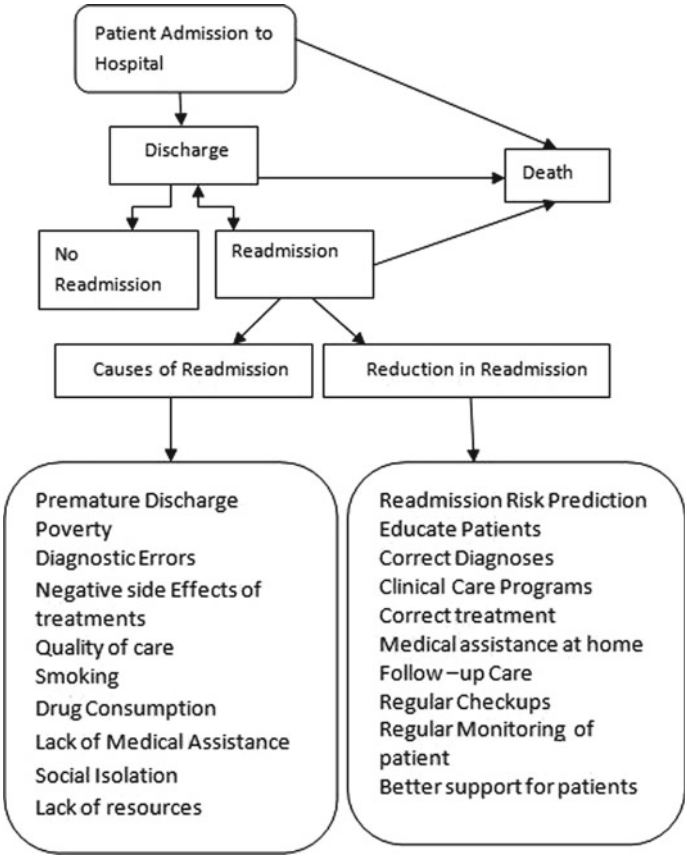



**Fig. 8** Re-admission process: causes and ways to reduce hospital re-admissions

## 5 Conclusion

Re-admission is the major cost driver for medicare systems. The explanations for why people return to the hospital soon after discharge are complicated. Unnecessary medical re-admissions reduce not just the quality of medical care, but it also results in high resource wastage in hospitals. Few attempts have been made to find out the appropriate time frame in forecasting re-admission. Factors such as poverty, gender, type of population, age, and seriousness of illness may lead to differences in the predicted time periods of the specified clinical conditions. Patients who are identified at high re-admission risk while hospitalization can aid to cut down re-admissions. Socioeconomic factors are mostly helpful in prediction, but usually, they are not present in the electronic medical record system of hospitals. Hence, even if there is a design of risk prediction models for a particular hospital, it does not offer outstanding performance.

Re-admissions influence the demographics of patients such as social, cultural, health, and financial factors; resource availability, care quality, and hospital bed. Hospitals treating underprivileged people and delivering inadequate treatment are the two major reasons for increased re-admissions. Another concern worth noting is that hospitals with fewer/no re-admission fines are more likely to be hospitals with adequate funds to devote their resources to improving the quality of treatment. Therefore, the penalization of hospitals with financial problems is not favorable. However, it is the responsibility of hospitals to strengthen patient care and adhere to follow-up services, but should never, at the same time, be prevented from helping the poorer and sicker populations. Larger, well-accomplished research should continue to gather evidence on the long-term efficacy of quality of treatment approaches.

## References

1. Al-Mardini M, Hajja A, Clover L, Olaleye D, Park Y, Paulson J, Xiao Y (2016) Reduction of hospital re-admissions through clustering based actionable knowledge mining. In: 2016 IEEE/WIC/ACM international conference on web intelligence (WI). IEEE, pp 444–448
2. Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A (2018) Predicting hospital re-admission via cost-sensitive deep learning. IEEE/ACM Trans Comput Biol Bioinform (TCBB) 15(6):1968–1978
3. Jencks SF, Williams MV, Coleman EA (2009) Rehospitalizations among patients in the medicare fee-for-service program. N Engl J Med 360:1418–1428
4. Medicare Payment Advisory Commission (MedPAC) (2007) Report to congress: promoting greater efficiency in medicare
5. Rao VR, Zolfaghar K, Hazel DK, Mandava V, Roy SB, Teredesai A Re-admissions score as a service (RaaS)
6. Yu S, Farooq F, Van Esbroeck A, Fung G, Anand V, Krishnapuram B (2015) Predicting re-admission risk with institution-specific prediction models. Artif Intell Med 65(2):89–96
7. Yu Z, Rouse WB (2017) A deeper look at the causes of hospital re-admissions. In: 2017 IEEE international conference on industrial engineering and engineering management (IEEM). IEEE, pp 919–923

8. Maddipatla RM, Hadzikadic M, Misra DP, Yao L (2015) 30 Day hospital re-admission analysis. In: 2015 IEEE international conference on big data (big data). IEEE, pp 2922–2924
9. (2012) Medicare to penalize 2217 hospitals for excess re-admissions KaiserHealth News. http://www.kaiserhealthnews.org/Stories/2012/August/13/medicarehospitals-re-admissions-penalties.aspx
10. Moskovitch R, Choi H, Hripcsak G, Tatonetti NP (2017) Prognosis of clinical outcomes with temporal patterns and experiences with one class feature selection. IEEE/ACM Trans Comput Biol Bioinform 14(3):555–563
11. Stojanovic J, Gligorijevic D, Radosavljevic V, Djuric N, Grbovic M, Obradovic Z (2017) Modeling healthcare quality via compact representations of electronic health records. IEEE/ACM Trans Comput Biol Bioinf 14(3):545–554
12. Zhang L, Liu H, Huang Y, Wang X, Chen Y, Meng J (2017) Cancer progression prediction using gene interaction regularized elastic net. IEEE/ACM Trans Comput Biol Bioinform (TCBB) 14(1):145–154
13. Zolfaghar K, Agarwal J, Sistla D, Chin S-C, Basu Roy S, Verbiest N (2013) Risk-o-meter: an intelligent clinical risk calculator. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1518–1521
14. Wang H, Wu J, Pan S, Zhang P, Chen L (2017) Towards large-scale social networks with online diffusion provenance detection. Comput Netw 114:154–166
15. Wang H, Wu J (2017) Boosting for real-time multivariate time series classification. In: AAAI, pp 4999–5000
16. Wang H, Zhang P, Tsang I, Chen L, Zhang C (2015) Defragging subgraph features for graph classification. In: Proceedings of the 24th ACM international on conference on information and knowledge management. ACM, pp 1687–1690
17. Wang H, Zhang P, Zhu X, Tsang IW-H, Chen L, Zhang C, Wu X (2017) Incremental subgraph feature selection for graph classification. IEEE Trans Knowl Data Eng 29(1):128–142
18. Kim S, Kim W, Park RW (2011) A comparison of intensive care unit mortality prediction models through the use of data mining techniques. Healthc Inform Res 17(4):232–243
19. Mao Y, Chen W, Chen Y, Lu C, Kollef M, Bailey T (2012) An integrated data mining approach to real-time clinical monitoring and deterioration warning. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1140–1148
20. Sushmita S, Khulbe G, Hasan A, Newman S, Ravindra P, Roy SB, De Cock M, Teredesai A (2016) Predicting 30-day risk and cost of "all-cause" hospital re-admissions. In: Workshops at the thirtieth AAAI conference on artificial intelligence
21. He D, Mathews SC, Kalloo AN, Hutfless S (2014) Mining high-dimensional administrative claims data to predict early hospital re-admissions. J Am Med Inform Assoc 21(2):272–279
22. Almayyan W (2016) Lymph diseases prediction using random forest and particle swarm optimization. J Intell Learn Syst Appl 8(03):51
23. Choudhry SA, Li J, Davis D, Erdmann C, Sikka R, Sutariya B (2013) A public-private partnership develops and externally validates a 30-day hospital re-admission risk prediction model. Online J Public Health Inform 5(2):219
24. Somanchi S, Adhikari S, Lin A, Eneva E, Ghani R (2015) Early prediction of cardiac arrest (code blue) using electronic medical records. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 2119–2126
25. Choi E, Schuetz A, Stewart WF, Sun J (2016) Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 112
26. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W (2016) Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. In: Advances in neural information processing systems, pp 3504–3512
27. Amarasingham R, Moore BJ, Tabak YP, Drazner MH, Clark CA, Zhang S, Reed WG, Swanson TS, Ma Y, Halm EA (2010) An automated model to identify heart failure patients at risk for 30-day re-admission or death using electronic medical record data. Med Care 48(11):981–988

28. Alluhaidan A, Lee E, Alnosayan N, Chatterjee S, Houston-Feenstra L, Dysinger W, Kagoda M (2015) Designing patient-centered mHealth technology intervention to reduce hospital re-admission for heart-failure patients. In: 2015 48th Hawaii international conference on system sciences. IEEE, pp 2886–2895
29. Khayyat A, Sequera C, Walk N, Wong E, Barbera J, Mazzuchi T, Santos J (2019) Decision support tool to estimate and reduce the probability of re-admission for congestive heart failure patients. In: 2019 Systems and information engineering design symposium (SIEDS). IEEE, pp 1–6
30. Vedomske MA, Brown DE, Harrison JH (2013) Random forests on ubiquitous data for heart failure 30-day re-admissions prediction. In: 2013 12th International conference on machine learning and applications, vol 2. IEEE, pp 415–421
31. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S (2011) Risk prediction models for hospital re-admission: a systematic review. JAMA 305(15):1688–1698
32. Facchinetti G, D'Angelo D, Piredda M, Petitti T, Matarese M, Oliveti A, De Marinis MG (2019) Continuity of care interventions for preventing hospital re-admission of older people with chronic diseases: a meta-analysis. Int J Nurs Stud 103396
33. Coleman EA, Min S-J, Chomiak A, Kramer AM (2006) A posthospital care transitions: patterns, complications, and risk identification. Health Serv Res 39(5):1449–1466
34. Demir E, Chaussalet T (2009) A systematic approach in defining re-admission. In: 22nd IEEE international symposium on computer-based medical systems. IEEE, pp 1–7
35. Yu K, Xie X (2019) Predicting hospital re-admission: a joint ensemble-learning model. IEEE J Biomed Health Inform
36. Baig MM, Hua N, Zhang E, Robinson R, Armstrong D, Whittaker R, Robinson T, Mirza F, Ullah E (2019) Machine learning-based risk of hospital re-admissions: predicting acute re-admissions within 30 days of discharge. In: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE, pp 2178–2181
37. Agarwal A, Baechle C, Behara R, Zhu X (2017) A natural language processing framework for assessing hospital re-admissions for patients with COPD. IEEE J Biomed Health Inform 22(2):588–596
38. Zeng L, Neogi S, Rogers J (2014) Statistical models for hospital re-admission prediction with application to chronic obstructive pulmonary disease (COPD) patients. In: Proceedings of the international conference on industrial engineering and operations management, Bali, Indonesia, pp 1–11
39. Hasan O et al (2010) Hospital re-admission in general medicine patients: a prediction model. J Gen Internal Med 25(3):211–219
40. Tran T, Luo W, Phung D, Gupta S, Rana S, Kennedy R, Larkins A, Venkatesh S (2014) A framework for feature extraction from hospital medical data with applications in risk prediction. BMC Bioinform 15(1):6596
41. Lee S, Wang S, Bain PA, Baker C, Kundinger T, Sommers C, Li J (2018) Reducing COPD re-admissions: a causal Bayesian network model. IEEE Robot Autom Lett 3(4):4046–4053
42. Lee S, Wang S, Bain PA, Kundinger T, Sommers C, Baker C, Li J (2018) Modeling and analysis of postdischarge intervention process to reduce COPD re-admissions. IEEE Trans Autom Sci Eng 16(1):21–34
43. Romeijn HE, Ahuja RK, Dempsey JF, Kumar A (2006) A new linear programming approach to radiation therapy treatment planning problems. Oper Res 54(2):201–216
44. Craft D (2007) Local beam angle optimization with linear programming andgradient search. Phys Med Biol 52(7):N127–N135
45. Earnshaw SR, Hicks K, Richter A, Honeycutt A (2007) A linear programming model for allocating HIV prevention funds with state agencies: a pilot study. Health Care Manage Sci 10(3):239–252
46. Joustra PE, de Wit J, Struben VMD, Overbeek BJH, Fockens P, Elkhuizen SG (2010) Reducing access times for an endoscopy department by an iterative combination of computer simulation and linear programming. Health Care Manage Sci 13(1):17–26

47. Arimond M, Vitta B, Martin-Prével Y, Moursi M, Dewey KG (2018) Local foods can meet micronutrient needs for women in urban Burkina Faso, but only if rarely consumed micronutrient-dense foods are included in daily diets: a linear programming exercise. Matern Child Nutr 14(1):
48. Hu P, Li S, Huang YA, Hu L (2019) Predicting hospital re-admission of diabetics using deep forest. In: 2019 IEEE international conference on healthcare informatics (ICHI). IEEE, pp 1–2
49. Duggal R, Shukla S, Chandra S, Shukla B, Khatri SK (2016) Predictive risk modelling for early hospital re-admission of patients with diabetes in India. Int J Diab Dev Ctries
50. Zolfaghar K, Meadem N, Teredesai A, Roy SB, Chin SC, Muckian B (2013) Big data solutions for predicting risk-of-re-admission for congestive heart failure patients. In: 2013 IEEE international conference on big data. IEEE, 64–71
51. Ross JS, Mulvey GK, Stauffer B, Patlolla V, Bernheim SM, Keenan PS, Krumholz HM (2010) Statistical models and patient predictors of re-admission for heart failure a systematic review. Health Serv Res 45(6):1815–1835
52. Wasfy JH, Singal G, O'Brien C, Blumenthal DM, Kennedy KF, Strom JB, Spertus JA, Mauri L, Normand SLT, Yeh RW (2015) Enhancing the prediction of 30-day re-admission after percutaneous coronary intervention using data extracted by querying of the electronic health record. Circ Cardiovasc Qual Outcomes 8(5):477–485
53. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S (2011) Clinician's corner risk prediction models for hospital re-admission a systematic review. JAMA 306(15):1688–1698
54. vanWalraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K, Austin PC, Forster AJ (2010) Derivation and validation of an index to predict early death or unplanned re-admission after discharge from hospital to the community. CMAJ 182(6)
55. Halfon P, Eggli Y, Pretre-Rohrbach I (2006) Validation of the potentially avoidable hospital re-admission rate as a routine indicator of the quality of hospital care. Med Care 44(11):972–981
56. Hasan O, Meltzer DO, Shaykevich SA, Bell CM, Kaboli PJ, Auerbach AD, Wetterneck TB, Arora VM, Zhang J, Schnipper JL (2010) Hospital re-admission in general medicine patients: a prediction model. J Gen Intern Med 25(3):211–219
57. Billings J, Dixon J, Mijanovich T, Wennberg D (2006) Case finding for patients at risk of re-admission to hospital: development of algorithm to identify high risk patients. BMJ 333
58. Hensher M, Edwards N, Stokes R (1999) International trends in the provision and utilisation of hospital care. BMJ 319:845–848
59. Capewell S (1996) Stemming the tide of re-admissions: patient, practice or practitioner? Br Med J 312:991–992
60. Evangelista L, Doering L, Dracup K (2018) Usefulness of a history of tobacco and alcohol use in predicting multiple heart failure re-admissions among veterans. Am J Cardiol. Accessed: 11-Oct-2018
61. Pierre-Louis B et al (2016) Clinical factors associated with early re-admission among acutely decompensated heart failure patients. Arch Med Sci. [Online]. Available: http://10.0.19.250/aoms.2016.59927. Accessed: 6-Mar-2019
62. Happ M, Naylor M, Roe-Prior P (1997) Factors contributing to rehospitalization of elderly patients with heart failure. J Cardiovasc Nurs. [Online]. Available:https://doi.org/10.1097/00005082-199707000-00008. Accessed: 11-Oct-2018
63. Amarasingham R et al (2010) An automated model to identify heart failure patients at risk for 30 day re-admission or death using electronic medical record data. Med Care J 48(11):981–988. [Online]
64. Kagabo R, Kim J, Zubieta JK, Kleinschmit K, Okuyemi K (2019) Association between smoking, and hospital re-admission among inpatients with psychiatric illness at an academic inpatient psychiatric facility, 2000–2015. Addict Behav Rep 9:
65. Nafis MT, Biswas R (2019) A secure technique for unstructured big data using clustering method. Int J Inf Technol. https://doi.org/10.1007/s41870-019-00278-x
66. Luthi C, Burnard B, McClellan M, Pitts R, Flanders D (2003) Is re-admission to hospital an indicator of poor process of care for patients with heart failure? Br Med J 13:46–51

67. Roland M, Dusheiko M, Gravelle H, Parker S (2005) Follow up of people aged 65 and over with a history of emergency admissions: analysis of routine admission data. BMJ 330:289–292
68. Hajja A, Touati H, Raś ZW, Studnicki J, Wieczorkowska AA (2014) Predicting negative side effects of surgeries through clustering. In: New frontiers in mining complex patterns. Springer International Publishing, pp 41–55
69. Frellick M (2015) Landmark report urges reform to avert diagnostic errors. Medscape
70. Mayr FB, Talisa VB, Balakumar V, Chang CCH, Fine M, Yende S (2017) Proportion and cost of unplanned 30-day re-admissions after sepsis compared with other medical conditions. JAMA 317(5):530–531
71. Nafis MT, Wazir S, Kumar A, Sharma DK (2020) Mining of high average utility itemset from interested items. Int J Sci Technol Res 9(4)
72. Paul S (2008) Hospital discharge education for patients with heart failure: what really works and what is the evidence? Crit Care Nurse J [Online]. Available: http://ccn.aacnjournals.org/content/28/2/66. Accessed: 12-Oct-2018
73. Rockwell JM, Riegel B (2001) Predictors of self-care in persons with heart failure. Heart Lung: J Acute Crit Care 30(1):18–25. https://doi.org/10.1067/mhl.2001.112503
74. Kutzleb J, Reiner D (2006) The impact of nursedirected patient education on quality of life and functional capacity in people with heart failure. J Am Acad Nurse Pract 18(3):116–123. https://doi.org/10.1111/j.1745-7599.2006.00107.x
75. Calvillo-King L, Arnold D, Eubank K, Lo M, Yunyongying P, Stieglitz H, Halm E (2018) Impact of social factors on risk of re-admission or mortality in pneumonia and heart failure: systematic review. J Gen Intern Med
76. Lindenfelf J, Albert N, Boehmer J et al (2010) HFSA 2010 comprehensive heart failure practice guideline. J Cardiac Fail 16(6):e1-e194
77. Konerman M, Hummel S (2014) Sodium restriction in heart failure: benefit or harm? Curr Treat Options Cardiovasc Med. [Online]
78. American cancer society guidelines for the early detection of cancer. http://www.cancer.org/healthy/findcancerearly/cancerscreeningguidelines/american-cancer-society-guidelines-for-the-early-detection-of-cancer
79. Raj JS, Ananthi JV (2019) Recurrent neural networks and nonlinear prediction in support vector machines. J Soft Comput Paradigm (JSCP) 1(01):33–40
80. Raj JS (2019) A comprehensive survey on the computational intelligence techniques and its applications. J ISMAC 1(03):147–159

# An Edge DNS Global Server Load Balancing for Load Balancing in Edge Computing


**P. Herbert Raj**

**Abstract** The information technology era faces an inflection spot due to current changes in the network industry due to the surrogation of cloud application implementation instead of data centre-based application implementation. The fiery growth of portable devices, sensor, and IoT-based devices would cause a stern effect on cloud application development and deployment in the network industry. The edge computing model is developed to enhance cloud application deployment. Edge computing is defined as where the computed calculations and storage of data are located in close physical proximity to improve the response time, reduce the latency, and improve the network bandwidth. This article analyses the ways to achieve the goals of edge computing with the assistance of load balancing.

**Keywords** Edge computing · Data centre · Global server load balancing (GSLB) · Disaster recovery plan (DRP) · Distributed denial of service (DDoS) · Internet of Things (IoT) · 5G · Latency

## 1 Introduction

This is an epoch of cloud computing. Currently, many people are using personal computers and desktops to access many of the centralized server computers. At the same time, IBM, Amazon, Microsoft, and Google are providing cloud services to their users to improve scalability, efficiency, reduce IT expenditure, 24 × 7 connectivity, and much more [1]. The cloud computing business could surpass $330 billion in 2020 due to the embracement of cloud computing [2]. This happened because users could access the resources from anywhere at any time. The upcoming 5G technology will enhance these capabilities because edge computing must provide high bandwidth to the connected users [3–5].

P. Herbert Raj (✉)
Bandar Seri Begawan, Brunei



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_57

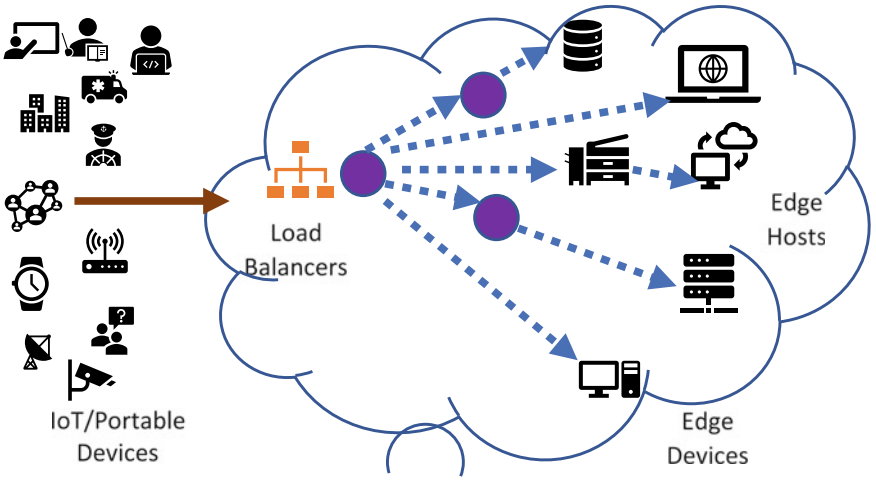



**Fig. 1** Edge architecture with load balancers

## *1.1 Need for Load Balancing in Edge Computing*

Edge computing tries to move the computations nearer to clients to minimize the delay and to reduce the bandwidth usage. The edge of a communication network is very close to the client; unlike the cloud network, the servers are placed in a faraway place. Cloud computing is a central network management system where applications are functioning in the data centres. Edge computing is also a central network management system but here applications are operating either in the device or in the network edge [6]. It intensifies the confidentiality and solitude of data being processed. Network instability or disruption will not affect the overall operations in edge computing [7]. Figure 1 depicts the edge computing architecture with load balancers.

## *1.2 Need for Load Balancing in Edge Computing*

Scalability is the key issue in load balancing. Some services may not be available to clients due to the network or website or server failure. This may lead to disparity in the network traffic distribution. Load balancing algorithms take a lead to resolve the unbalanced traffic distribution. Even though numerous algorithms are available, achieving optimal load balancing is an ultimatum. The cloud bursting method is used in some servers to settle the load balancing issues in the edge environment [8]. The application of load balancing is greatly influenced by the tendencies of edge computing. The job of the load balancer is to migrate together with other apps. Data centres of edge have load balancers that utilize the global server load balancing

(GSLB) for linking the different edge data centres. This will maintain an effective reply time by spreading the load over multiple edge servers [9].

### *1.3 Load Distribution*

Load in the network will be distributed evenly between clients and servers by using various load balancing algorithms. The algorithm will be selected based on what sort of facility provided, and the present stage of the network will be taken into consideration. There are few load balancing algorithms used to balance the load, namely round-robin, weighted least connection, and resource-based.

### *1.4 Edge Computing and Load Balancing*

Internet of Things can able to nexus with numerous portable and smart gadgets and automatically turn out to be a component of varying cutting-edge applications. This produces a huge volume of data, and it may increase in future [10, 11]. The data centres of clouds are positioned very in far-flung locations from these gadget users, and the bandwidth is limited [12]. So, latency in request-reply is inevitable. Edge computing assists to reduce this distance and moving closely from federal services to the boundary of the network. These edge devices may handle a huge sum of data and concurrent tasks without any latency [13]. Nevertheless, this might lead to unbalanced traffic distribution due to the huge amount of data to be processed. Multiple data centres in the edge can be combined as a single virtual one. So, load balancing can be done here by using GSLB. Nowadays, edge computing load balancing is drawn the attention of scholars because of its physical proximity to the users and latency reduction [1]. In this article, the process of GSLB comfort edge load balancing is discussed in detail.

## 2 GSLB Load Balancing

In GSLB, the network traffic is disseminated to the numerous linked servers in this universe. This is one of the most credible load balancing methods for industries, and it greatly minimizes latency. GSLB empowers multi-edge data centres to control and manipulate the availability of the resources placed across the world. This can be achieved by incorporating virtual multi-edge data centres, several servers, and a cloud environment. It has optimized traffic redirection methods in case of any failure or service intrusion over the network distribution. The user's request is sent to a location where it can obtain the best service and also examines the availability of the resources.

### *2.1 Need for Edge DNS GSLB*

(a) **Load Balancers**: Load balancers function well on a smaller scale, which means when the end users' devices and servers are limited. LBs may not function well in the multi-cloud with numerous data centre architectures due to their complicated topology. For the LBs, there is a need for definite design and bandwidth to link the data centres and user devices. It is not possible to provide these facilities for all the services. To alleviate this problem, DNS can be used [14]. Here, user requests can select one of the offered IP address at random and send their traffic to a particular destination. This will disseminate the load amid data centres. The LBs mounted in the local data centres lever the remaining traffic.

(b) **DNS**: DNS disseminates the network load to numerous addresses with a similar application label. Setting up a DNS connection is very straightforward, and the service provided by DNS can be easily used by users. In load balancing, DNS redundancy cannot be updated automatically when a server is added or removed. Some automatic updates or manual updates are needed at this point. This may lead to application unavailability for some clients. Global DNS system may provide a similar reply to all the connected users. Some may need a different reply according to their site location. So vibrant updates and site-specific replies are mandatory here. To alleviate this sort of problem, DNS global is hosted [14].

(c) **DNS GSLB**: DNS GSLB is able to reply vibrantly to the client requests and robotically updates the server. The disadvantage of this method is that those who need to use this service have to be in this geographical zone. This system is highly centralized. Always on service availability may not be possible due to the capability to find out the available servers. This problem leads to find another better method to route the traffic smartly and directly.

(d) **Edge DNS GSLB**: Edge DNS GSLB is considered to be the best solution for traffic routing because it offers the best routing decision to reach the destination for a requested client. It can answer any client's request without any centralized committed zone. The traffic routing of edge DNS GSLB can be integrated with characteristics of a recursive DNS server which offers a lot of benefits to the clients. The benefits of this method are scalability, swift responsiveness, obtainability of local data, proficient caching, security, optimized DNS traffic, and augments multicentre robustness [14].

## 3 Edge DNS GSLB and Multi-cloud

Edge DNS GSLB produces a professional and modest way to load balance the network load. The application routing choice will be made at the location of users which is closer to the edge network. The forthcoming sessions describe the advantages of executing Edge DNS GSLB.

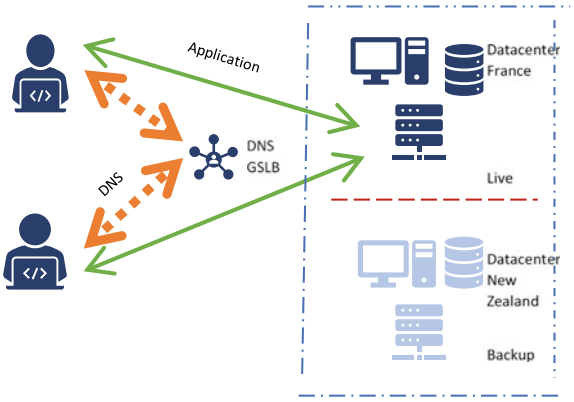



**Fig. 2** Live connection. *Source* https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf

### *3.1 Advanced Disaster Recovery*

Industries host significant applications in a particular server that will be available in the disaster recovery plan sites. These kinds of applications will be accessed and used by numerous users. IP address maintenance will be handled by DNS. In case of any failure arises, these applications will not be accessible. The new location of the network must be reflected in the IP address. Without blunders, changing the DNS information is impossible. Moreover, it is a time laborious task. All users switching over to the DRP site would complicate things furthermore.

At the application level, Edge DNS GSLB permits amalgamating manifold design facilities and switching over.

(a) **Failovers**: Here, each application is established with two nodes, one is to link the key site and another one is to link the reserve site. If the key site is incapacitated, then all the applications are connected to the DR site. This will be done automatically [14]. Figure 2 shows the live connection between the user and the application data centres.

### *3.2 Sharing of Network Load Between Multiple Data Centres*

Consider a scenario, an application is accommodated in several data centres for distributing the network load. The users' requests will be directed according to their own site to the desired data centres. The challenge is how to make this possible for a user-specific site. Figure 3 describes the traffic routing with a health check.

(a) **Dynamic Distribution**: The level of the application server link will be checked on an ordered basis by Edge DNS GSLB to determine whether the users can be pointed to it. This will decide on which is the best link for the user [14].

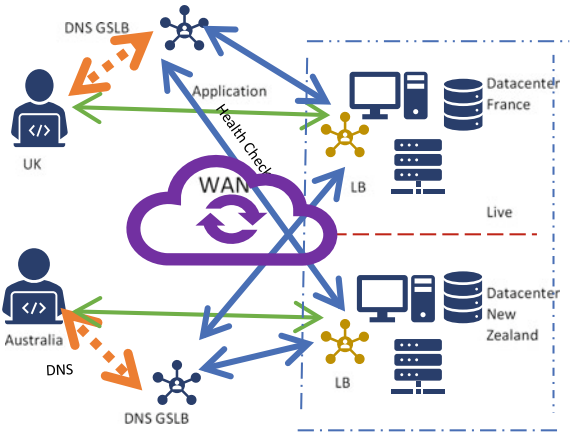



**Fig. 3** Application traffic routing with a health check. *Source* https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf

### *3.3 WAN Failure Detection*

In the case of WAN failure, clients' requests should be led to the application server which is accessible and the top pick. The challenge is that accessing from a middle of a server to a client placed in a far-flung site may not be possible all the time. Edge DNS GSLB identifies any differences in the network among the far-flung location and the application server and gives the best solution for routing [15]. Figure 4 shows the selection of the best destination routing.

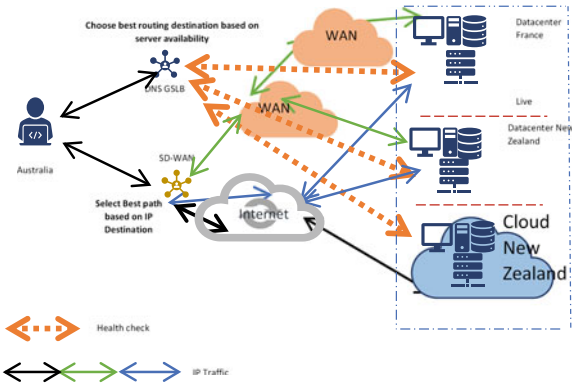



**Fig. 4** Selecting the best destination. *Source* https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf

## 4 Conclusion

Until now, IP address resolution is done by DNS service. This is a significant procedure in IP networks because it permits users to access the application server. The operational requirements made the network engineers devise new protocols and devices to enable the users to access everything smoothly and securely. When data centres provide services to the users to store or retrieve data and a lot of security measures are devised to access the data securely. The design complications are diminished with the help of Edge DNS GSLB. Load balancing and multi-cloud management can be designed undoubtedly with the help of Edge DNS GSLB. The significant benefits of Edge DNS GSLB have increased scalability, dexterity, robust multi-cloud management, user experience, and simple DR plans. Most importantly, this load balancing technique minimizes the energy consumptions of the processor without lowering the edge network functioning. Finally, this method guarantees the application accessibility for the clients, if any WAN breakdown happens. The implementation benefits of Edge DNS GSLB perfect match for load balancers, SD-WAN, and application delivery controllers to make correct routing decisions. This is very straightforward to implement and expands huge benefits.

## References

1. Li G, Yao Y, Wu J, Liu X, Sheng X, Lin Q (2020) A new load balancing strategy by task allocation in edge computing based on intermediary nodes. EURASIP J Wireless Commun Networking 2020(3). https://doi.org/10.1186/s13638-019-1624-9
2. Nick Galov (2020) 25 must-know cloud computing statistics in 2020, hosting tribunal. https://hostingtribunal.com/blog/cloud-computing-statistics/
3. Ravi Kumar P, Herbert Raj P, Jelciana P (2017) Exploring security issues and solutions in cloud computing services. Cybern Inf Technol 17(4):29. Print ISSN: 1311-9702; Online ISSN: 1314-4081
4. Abdulmohson A, Pelluri S, Sirandas R (2015) Energy efficient load balancing of virtual machines in cloud environments. Int J Cloud-Comput Super-Comput 2(1):21–34. http://dx.doi.org/10.21742/ijcs.2015.2.1.03
5. Efficient IP (2019) How edge DNS GSLB ensures app availability during WAN failure. https://www.efficientip.com/dns-gslb-wan-failure/
6. Cloud Fare (2020) What is edge computing. https://www.cloudflare.com/learning/serverless/glossary/what-is-edge-computing/
7. Gonzalez J, Hunt J, Thomas M, Anderson R, Mangla U (2020) Edge computing architecture and use cases. LF Edge Premium Member Company IBM. https://www.lfedge.org/2020/03/05/edge-computing-architecture-and-use-cases/
8. Nugara A (2019) Load balancing in Microsoft Azure. O'Reilly Publisher(s) Inc., ISBN: 9781492053927
9. Yue F (2020) Edge computing and load balancing. Kemp technologies. https://kemptechnologies.com/blog/edge-computing-and-load-balancing/
10. Wan S, Zhao Y, Wang T, Gu Z, Abbasi QH, Choo KKR (2019) Multi-dimensional data indexing and range query processing via Voronoi diagram for Internet of things. Future Gener Comput Syst 91:382–391

11. Zaslavsky A, Perera C, Georgakopoulos D (2013) Sensing as a service and big data. arXiv preprint arXiv:1301.0159
12. Herbert Raj P, Ravi Kumar P, Jelciana P (2016) Mobile cloud computing: a survey on challenges and issues. Int J Comput Sci Inf Secur (IJCSIS) 14(12). ISSN 1947–5500
13. Zhu T, Shi T, Li J, Cai Z, Zhou X (2018) Task scheduling in deadline-aware mobile edge computing systems. IEEE Internet Things J 1. https://doi.org/10.1109/jiot.2018.2874954
14. Efficient IP (2020) Edge DNS GSLB complement your load balancing and multi-cloud strategy. EfficientIP, SAS
15. Efficient IP (2020) Edge DNS GSLB use cases improving UX, DRP and datacenter agility. EfficientIP, SAS. https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf

# Network Intrusion Detection Using Cross-Bagging-Based Stacking Model


**S. Sathiya Devi and R. Rajakumar**

**Abstract** Network-based information transmission has brought huge convenience for users in terms of the ease of use. However, the increased transaction not only lures fraudsters but also makes attack detection with a complicated process and mandates scalable models that can handle big data. This paper presents a network intrusion detection model that uses a hybrid ensemble model to identify intrusions in network transmission process. This work proposes the cross-bagging-based stacked ensemble model, which is a two-layered prediction mechanism used for operating on the complex network data. The first layer that contains a modified bagging mechanism is called the cross-bagging, where the results are passed to the second layer for final prediction. Experiments were conducted by using the NSL-KDD dataset. The usage of ensemble modelling enables the model to be effectively parallelized and also ensures high scalability of the model. This ensures effective prediction even on data with large volumes and high velocity. Comparisons with recent models show high performance for the proposed model.

**Keywords** Network intrusion detection · Cross-bagging · Stacking · Anomaly detection · Ensemble model

## 1 Introduction

Increased usage of networks and network-based transactions in recent times is mainly due to the ease of use. However, the flexibility comes with the vulnerability associated with it [1]. Packets transmitted in networks are prone to attacks by creating a huge security risk in the transmission process. Although this cannot be avoided altogether, it can to a large extent be detected and prevented at the earliest [2, 3]. Network intrusion detection systems play a pivotal role in performing this process effectively. Faster detection process results in a better system by providing elevated levels of security. Further, the large amount of data created in the current scenario makes

S. Sathiya Devi · R. Rajakumar (✉)
BIT Campus, Anna University, Tiruchirappalli, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_58

intrusion detection in networks as a significant challenge. This further results in the requirement of big data processing techniques for the detection process.

Technically, intrusion detection systems can be categorized into two; they are intrusion detection systems (IDS) and intrusion prevention systems (IPS) [4]. IPS is a proactive model, while IDS is a reactive model. IDS is a more adopted problem when compared to IPS [5]. This work presents an IDS model called the cross-bagging-based stacked ensemble (CBSE). Network data is complex, hence requires a complex model for prediction. The proposed CBSE model is composed of two levels; the begging process and the stacking-based prediction. This enables the model to create higher-level abstractions, hence provides better predictions.

## 2 Literature Review

Intrusion detection plays a pivotal role in networks paradigm with the effective network functioning parameters. IDS systems are categorized based on their operational nature. This section categorizes the IDS models into three categories; they are hybrid/ensemble models, machine learning-based models and ensemble models.

A hybrid model for detecting intrusions in network data was presented by Aljawarneh et al. [6]. This model is built in two layers; the first layer is composed of vote algorithm based on the information gain, and the next layer is composed of multiple machine learning models for prediction. Another hybrid technique that combines multiple models for prediction is presented by Ravale et al. [7]. A bagging-based intrusion detection model was proposed by Gaikwad et al. [2]. It uses partial decision trees for the prediction process. A context-aware IDS model for real-time environments was proposed by Pan et al. [8]. This model is designed for automated operations on networks. Other hybrid models include unsupervised model by Carrasco et al. [9], reinforcement learning-based model by Caminero et al. [10], deep learning-based model by Gurung et al. [11] and fuzzy cognitive maps (FCM)-based model by Chen et al. [12].

A support vector-based model that uses both classification and regression for prediction was presented by Bamakan et al. [13]. The model has been designed to operate on data imbalance and the skewed distribution of the data. Several other models utilizing support vector machine (SVM) for the prediction process are incremental SVM model by Yi et al. [14], principle component analysis (PCA)-based SVM model by Kuang et al. [15] and feature selection-based SVM model by Ahmad et al. [16]. An IDS model using lazy Learning was presented by Chellam et al. [17]. This model aims to reduce the computational complexity of the detection mechanism. An extreme learning machine-based model that uses incremental learning as the learning strategy was presented by Wang et al. [18].

A genetic algorithm-based IDS model was proposed by Pawar et al. [3]. It uses chromosomes of varying lengths for the prediction process. A combination of meta-heuristic models is also used in literature, creating hybrid models. An artificial bee colony (ABC) and AdaBoost-based IDS model was presented by Mazini et al. [19].

ABC is used for feature selection and AdaBoost is used for the detection process. A linear genetic programming-based IDS model was proposed by Hasani et al. [20]. It also incorporates the bees algorithm in its prediction process. Other swarm optimization models include articles by Gupta and Shrivastava [21], Sujitha and Kavitha [22], and firefly-based model by Selvakumar and Muneeswaran [23].

## 3 Cross-Bagging-Based Stacked Ensemble (CBSE)

Intrusion detection in networks has become a major concern due to the sensitivity of the information transmitted via them and also because of the increased attacks in network by fraudsters. This work proposes an ensemble-based model, cross-bagging-based stacked ensemble (CBSE) that aims to provide a complex analysis architecture that detects network intrusions significantly faster than the conventional models.

The proposed CBSE model is composed of three major phases; the initial preprocessing phase prepares the data for predictions, the cross-bagging ensemble modelling phase reads the data and prepares the decision rules for predictions, the final stacking phase operates on the predictions from the previous phase to determine the final predictions. Algorithm for the proposed model is given below.

**Algorithm**

1. Pre-process training data
2. Data segregation to form training and testing data
3. Divide training data into ten distinct subsets
4. Create base learners (Decision Tree and Random Forest)
5. For every subset s created (except for the last):
   (a) Pass to a distinct base learner for training
6. Use the last subset for prediction
7. Append prediction to the result set
8. Shuffle training data
9. Repeat steps 3–8 until required number of prediction results are obtained
10. Create training data (second level) for the stacking phase using the predictions and actual class
11. Train logistic regression model using the second-level training data
12. Pass predictions from this as the final predictions.

## 4 Data Preprocessing and Segregation

Network data in general is composed of several network attributes. Several of these properties represent details about the source, destination and some standard attributes

that are to be inserted into the network packets. These attributes however might not be needed for the machine learning model.

Data preprocessing is performed by identifying categorical and string attributes and converting them to numerical formats. Numerical features are the only type of features that can be processed by machine learning algorithms. Hence, categorical attributes are encoded to form numerical attributes. This work uses one-hot encoding method to convert categorical data to numerical formats. String data corresponds to data that cannot be encoded due to the generic (like Address) or very specific nature of data (like ID). Hence, they are eliminated from the data. This forms the end of the data preparation phase. The data is then segregated into training and test data. The division is performed in the ratio 7:3, where 70% of data is used for training and 30% of data is used for testing.

## 5 Cross-Bagging Ensemble Creation

Ensemble modelling is the process of using multiple machine learning models to perform predictions rather than relying on a single model. Ensemble models are considered to operate best when data involved for prediction is complex. Several ensemble modelling techniques are available in literature, namely bagging, boosting, stacking, etc. This work uses a combination of bagging and stacking models for prediction. This phase creates a bagging model for the prediction process.

Bagging or bootstrap aggregation is the process of creating various subsets of the data called bags and training multiple machine learning models based on the created bags. These models are used for prediction and the predictions are aggregated using a combiner to provide the final results. The proposed model modifies the bagging process by varying the bag creation mechanism and by replacing the combiner phase with the stacking process.

The proposed bag creation process is performed by splitting the data into ten distinct subsets. Overlaps between the subsets are eliminated in order to constrain the training data within a specific size. The number of subsets to be divided depends on the training data and the data similarity levels. This is determined by the domain expert. Nine of the generated subsets are used for training and the final set is used for prediction. The resultant predictions are recorded for processing by the stacking model. The dataset is then shuffled to rearrange the data instances. The shuffled dataset is again split into ten segments, nine of which are used for training and one for prediction. The predicted results are appended with the previous prediction set. The process of shuffling, splitting and prediction is repeated multiple times, and the results are iteratively appended to the prediction set. This process is repeated until sufficient amount of data is obtained for the next level stacking process.

Predictions are performed by multiple heterogeneous classifiers, rather than a single model. This aims to improve the prediction level of the data. This work uses decision tree and random forest for prediction. Although both are tree-based models, decision tree is a weak classifier, while random forest is a strong classifier. Being

weak classifiers, decision trees have the capability to handle dynamic data. Since the domain requires dynamism, decision trees are used as a part of the ensemble. Random forest, being a strong ensemble, can effectively overcome the overtraining issue experienced by the decision tree model. Predictions of both these models are combined to form the input data for the stacking model.

## 6 Stacking Phase for Final Prediction

The predicted results obtained from the previous phase are combined with the actual results to form the training data for the stacked model. Stacking is the process of operating on predictions of a model rather than on the actual data to determine the final predictions. Stacking operates to determine the predicting nature of a model. The model creates rules based on the predicting behaviour of models in the previous phase, rather than creating rules from the training data. This provides a two-level abstraction, hence makes the model effective even on highly complex data.

Logistic regression is used as the model of choice for the stacking phase. Although being a primitive model, logistic regression was observed to operate effectively on less complex data to provide highly effective predictions. The low-computational complexity exhibited by the CBSE model makes it a suitable choice for the second-level prediction.

The test data is passed to decision tree and random forest and the predictions are aggregated and passed to the stacked model. The predictions from stacked model are passed as the final predictions.

## 7 Results and Discussion

The cross-bagging-based stacked ensemble (CBSE) model has been implemented in Python. NSL-KDD is used to validate the model performance [24]. ROC curve for the CBSE model is shown in Fig. 1. ROC curve creates plots between TPR and FPR values. High TPR and low FPR are indicative of an effective classifier. The CBSE model shows TPR of 0.99 and very low FPR (<0.1), indicating good performance of the CBSE model.

PR curve for the CBSE model is shown in Fig. 2. PR curve is plotted with precision and recall in the axes. Precision represents the effectiveness of the classifier in correctly identifying the intrusions, while recall represents the capability of the model in labelling intrusions from all of the test instances. Both are expected to be high in a good model. High precision represents that most of the positively labelled instances are actually positive and high recall represents that the model has correctly labelled most of the positive instances as positive. CBSE model exhibits high precision and recall showing its prediction efficiency.

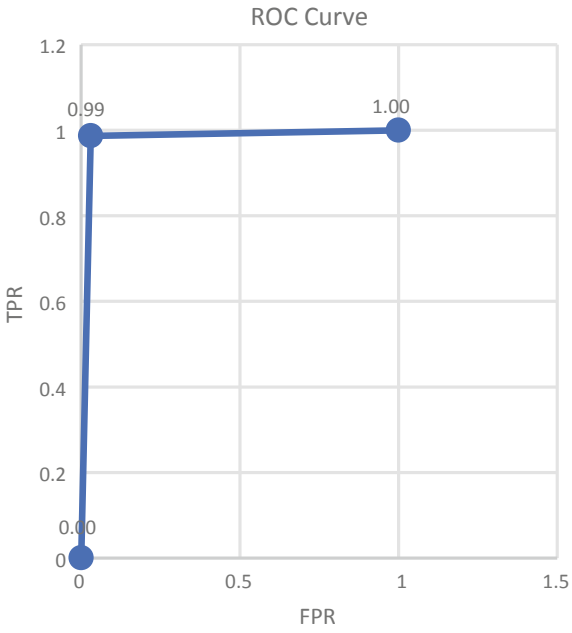



**Fig. 1** ROC curve

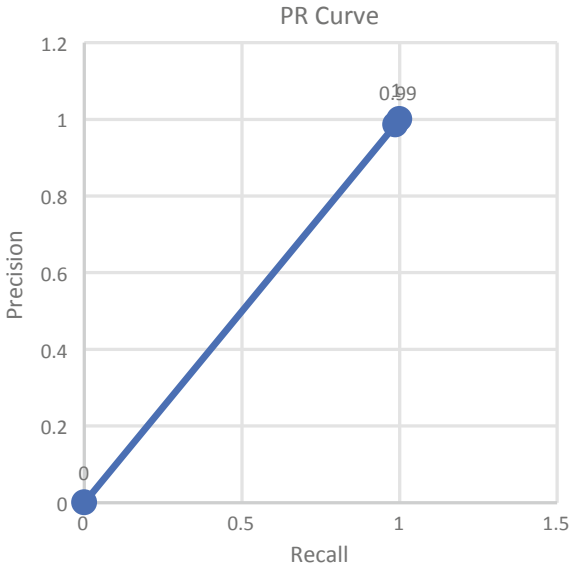



**Fig. 2** PR curve

Aggregate metrics indicate an overall view of the classifier performance. This work considers accuracy, precision and F1-score for analysis (Fig. 3). A comparison is also performed between the CBSE model, RampLoss model [13] and deep learning-based model [11]. RampLoss is the recent state-of-the-art model for intrusion detection in networked environments. The CBSE model exhibits either equivalent (accuracy at 99%) or higher performance (improvement of 1% on F1-score and 2% on precision) in terms of all the metrics. The high prediction capability of the CBSE model is explicit in the comparison. Although the predictions are similar, the ensemble model is parallelizable in nature and hence exhibits high scalability levels, making the model effective in big data environments.

The performance of CBSE is tabulated and shown in Table 1. The metrics show high performance (>97%) in terms of all true prediction metrics and (<3%) in terms of all the false predictions. This indicates high performance of CBSE.

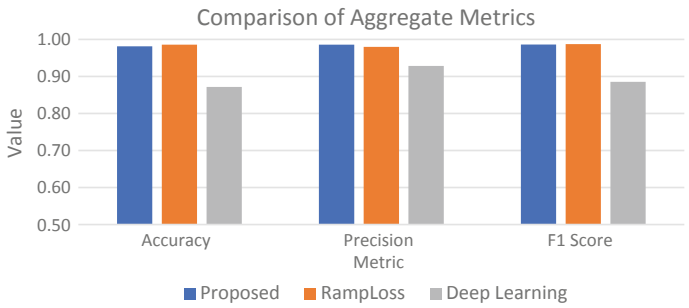



**Fig. 3** Performance comparison of aggregate metrics with RampLoss and deep learning models

**Table 1** Performance of CBSE model

| Metric | Value |
|---|---|
| FPR | 0.03 |
| TPR | 0.99 |
| Recall | 0.99 |
| Precision | 0.99 |
| F1-score | 0.99 |
| Accuracy | 0.98 |
| TNR | 0.97 |
| FNR | 0.01 |

## 8 Conclusion

Detecting network intrusions at the earliest enables secure network environments, hence encourages users to perform even highly sensitive transactions. This work presents an effective intrusion detection model, CBSE, to perform intrusion detection. The model is composed of two layers; the cross-bagging layer and the stacking layer. Experimental results and comparisons indicate high performances by the CBSE model.

Novelty of the model is based on the usage of varied combination of training and test sets in the cross-bagging layer results in forming rules that can handle the complex nature of the data under analysis. General bagging models are homogeneous in nature. The heterogeneity introduced in the process results in handling data insufficiency and also avoids overtraining. The stacking layer enables handling the complexity of the input data with the secondary abstraction. This high efficiency makes the proposed CBSE model highly effective in terms of predictions in real time.

## References

1. De la Hoz E, De La Hoz E, Ortiz A, Ortega J, Prieto B (2015) PCA filtering and probabilistic SOM for network intrusion detection. Neurocomputing 164:71–81
2. Gaikwad DP, Thool RC (2015) Intrusion detection system using bagging with partial decision treebase classifier. Procedia Comput Sci 49:92–98
3. Pawar SN, Bichkar RS (2015) Genetic algorithm with variable length chromosomes for network intrusion detection. Int J Autom Comput 12(3):337–342
4. Upasani N, Om H (2019) A modified neuro-fuzzy classifier and its parallel implementation on modern GPUs for real time intrusion detection. Appl Soft Comput 105595
5. Selvakumar K, Karuppiah M, SaiRamesh L, Islam SH, Hassan MM, Fortino G, Choo KKR (2019) Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in WSNs. Inf Sci 497:77–90
6. Aljawarneh S, Aldwairi M, Yassein MB (2018) Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. J Comput Sci 25:152–160
7. Ravale U, Marathe N, Padiya P (2015) Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function. Procedia Comput Sci 45:428–435
8. Pan Z, Hariri S, Pacheco J (2019) Context aware intrusion detection of building automation systems. Comput Secur. https://doi.org/10.1016/j.cose.2019.04.011
9. Carrasco RSM, Sicilia MA (2018) Unsupervised intrusion detection through skip-gram models of network behavior. Comput Secur 78:187–197
10. Caminero G, Lopez-Martin M, Carro B (2019) Adversarial environment reinforcement learning algorithm for intrusion detection. Comput Netw
11. Gurung S, Ghose MK, Subedi A (2019) Deep learning approach on network intrusion detection system using NSL-KDD dataset. Int J Comput Netw Inf Secur (IJCNIS) 11(3):8–14
12. Chen M, Wang N, Zhou H, Chen Y (2018) FCM technique for efficient intrusion detection system for wireless networks in cloud environment. Comput Electr Eng 71:978–987
13. Bamakan SMH, Wang H, Shi Y (2017) Ramp loss K-support vector classification-regression; a robust and sparse multi-class approach to the intrusion detection problem. Knowl-Based Syst 126:113–126

14. Yi Y, Wu J, Xu W (2011) Incremental SVM based on reserved set for network intrusion detection. Expert Syst Appl 38(6):7698–7707
15. Kuang F, Xu W, Zhang S (2014) A novel hybrid KPCA and SVM with GA model for intrusion detection. Appl Soft Comput 18:178–184
16. Ahmad I, Hussain M, Alghamdi A, Alelaiwi A (2014) Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components. Neural Comput Appl 24(7–8):1671–1682
17. Chellam A, Ramanathan L, Ramani S (2018) Intrusion detection in computer networks using lazy learning algorithm. Procedia Comput Sci 132:928–936
18. Wang CR, Xu RF, Lee SJ, Lee CH (2018) Network intrusion detection using equality constrained-optimization-based extreme learning machines. Knowl-Based Syst 147:68–80
19. Mazini M, Shirazi B, Mahdavi I (2018) Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms. J King Saud Univ-Comput Inf Sci
20. Hasani SR, Othman ZH, Mousavi Kahaki SM (2014) Hybrid feature selection algorithm for intrusion detection system. J Comput Sci 10:1015–1025
21. Gupta M, Shrivastava SK (2015) Intrusion detection system based on SVM and bee colony. Int J Comput Appl 111:27–32
22. Sujitha B, Kavitha V (2015) Layered approach for intrusion detection using multiobjective particle swarm optimization. Int J Appl Eng Res 10:31999–32014
23. Selvakumar B, Muneeswaran K (2019) Firefly algorithm based feature selection for network intrusion detection. Comput Secur 81:148–155
24. Ring M, Wunderlich S, Scheuring D, Landes D, Hotho A (2019) A survey of network-based intrusion detection data sets. Comput Secur

# Enterprise Network: Security Enhancement and Policy Management Using Next-Generation Firewall (NGFW)


**Md. Taslim Arefin, Md. Raihan Uddin, Nawshad Ahmad Evan, and Md Raiyan Alam**

**Abstract** Network security is considered as a major task in network architecture. A network administrator had to focus, and it is defined and demonstrated as the rules, plans, and procedures followed by a network administrator to protect the network devices from different threats, and simultaneously, the passive and active attacks are generated from various vulnerable sources. Further, the unauthorized users must be prevented from accessing the network. There are different types of threats that need to be identified, explored, and take a step for preventing it, wherein the attacks are like DoS and DDos attracts, Aurora attacks, malware attack, port scanning, password sniffer, IP spoofing, session hijacking, and man-in-the-middle attacks, and cyber-attacks. This could be done with the help of firewalls, which can secure the network from malicious attacks. This paper is more focused on strong policy and performs incredible directions for averting the mentioned attacks. Firewalls are one of the strongest hardware attachments to secure the zone of networking sectors like local large, multinational, or enterprise networks. The deployment of firewalls that enforce an organization's security policy is network devices. For this kind of tiresomeness, the concern of this paper is to enhance and develop network security like IPsec VPN, strong masquerades, port forwarding, create a trusted zone on WAN and LAN side, etc., based on the firewall by the execution of various tasks and different policies.

**Keywords** Network security · NGFW · Firewall · Enterprise network

Md. T. Arefin (✉) · Md. R. Uddin · N. A. Evan
Daffodil International University, Dhaka, Bangladesh
e-mail: arefin@diu.edu.bd

M. R. Alam
Texas A&M University-Kingsville, Kingsville, USA
e-mail: md_raiyan.alam@students.tamuk.edu



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_59

## 1 Introduction

Now, in this modern world, computer network system increases rapidly day by day. Therefore, a strong security system is very much important to protect the network. In a computer network, users interface are threatened from all kind of attacks from hackers. So, the development of network systems is the most required step where unwanted attacks are rising significantly. A large number of these dangers have turned out to be cunningly practiced assaults causing harm or submitting robbery. Moreover, government and business are massively dependent on the use of the Internet, and there are numerous advantages for using of Internet. So, ensuring secure applications for this sector to prevent the risk. Firewall concepts and technologies are nowadays a major topic that should be a part of any information security curriculum. Firewalls are security devices or software tools that are used to apply by implementing security rules to filter inward and outward network traffic of an organization's security policy. However, commercial firewalls are nowadays designed to be used by professionals and are not usually very appropriate for the academic environment. That is the most commercial firewalls do not allow educators to suggest advanced hands-on lab exercises on firewall concepts and technologies and offer usually complex user interfaces to con the firewalls and manipulate the filtering rules. For example, [1] low-level filtering rules are not to be allowed in commercial firewalls by manipulating the six flags of TCP, [2] also do not allow the lack of the ability to deal with the efficiency and the consistency of the filtering rules, [3] do not allow managing and viewing the table of stateful sessions, do not offer means to enhance the firewall packet processing performance by reordering the orders of the filtering rules and their rule-fields or using early packet rejection and acceptance mechanisms [4, 5] do not allow implementing rules to filter the application layers' contents (known as deep packet inspection (DPI)) and are unable to deal with common denial of service (DoS) attacks targeting the firewalls [6]. Consequently, commercial firewalls do not allow both the instructors to offer advanced hands-on lab exercises on firewalls, and the students to better anatomize advanced firewall concepts and technologies. An enterprise network is a network that helps to connect associate's PCs and related gadgets crosswise over offices and work group systems. An enterprise network decreases correspondence conventions, encouraging framework, and gadget interoperability and also enhanced the interior and outside enterprise data management which is called a corporate network. An enterprise network is to eliminate different clients and work groups as the key motivation. All methods should be able to communicate with each other and provide and recover data. Also, physical processes and devices should be able to maintain and give favorable reliability, performance, and security. Enterprise computing models are developed of established enterprise communication protocols and strategies for this purpose, facilitating exploration and improvement [7]. On the other hand, an enterprise network includes the LAN/WAN based on departmental and operational supplies [7]. Since the improper security model of an enterprise network exists, this paper gives an overview of the different kinds of security attacks of an enterprise network and how to prevent these attacks. And, finally, gives a properly deployed security The objectives of this work are to publish and define the

idea of attack and threat to a computer network, to highlight different allaying techniques used to complicate threats and attacks, to illustrate the procedure to implement the best security model for an enterprise network [8]. The various types of attack and different kinds of mitigation techniques are discussed in this paper. Finally, it implemented a secure network which prevents those attack. Finally, this paper will highlight on the importance of network security and also figure out the improper configuration of security in an enterprise network and, also, explain the newly deployed security model of an enterprise network. This newly deployed security model minimizes threats to give a better secure network. Also, this security-enhanced model can be resolved the future attack.

## 2 Related Work

A network admin can control the authorization of access in a network by deploying network security. To prevent the attack and ensure a safe and stable network, it is mandatory to exploit security. It can be easily maintained what services are allowed and what is not by using security policies in a firewall. Using a firewall for network security can be a sophisticated way to secure the service ports from the attackers, maintain valid user access for valid services, inspect the traffic, and detect the harmful contents like malware, Trojans, and other computer worms [7, 8]. Nowadays, security specialists are using Wireshark to capture packets and taking logs for analyzing the network system. Different kinds of encryption methods and security algorithms are used for the host to host communication. The communication technology is making the world smaller by enlarging and interconnecting the IP network. People from every stage are now adopting network security for creating a safe network environment of their data. The network security is now getting more priority for the scalable service, and the research goes to analyze it [8]. To ensure a better security environment, it is important to have security in all layers of the network. There are seven layers in the open system interconnection (OSI) model, and it is required to ensure security in all layers. In general, people think about only workstation security by applying security features in the application layer. To protect the data, it is important to exploit the security feature in all layers [9]. There can be any kind of malicious attack happen from any source when any workstation gets connected with the Internet. There can be two kinds of attacks in general, and those are ‘active attack’ and ‘passive attack.’ In the active attack, the attacker will ambush the target directly and will try to modify the system or information. In another word, the passive attack will make the attacker breach the transit security of the network to analyze user data and afterward monitor. To avoid these, a network security engineer must apply some strong security framework [10]. As a next-generation firewall, ‘Fortinet’ has a tremendous contribution by delivering high performance, sophisticated security solutions, and integrated features. This is more capable to analyze the vulnerable traffic and a better option for the security engineers [11]. Security architecture added the feature like intrusion prevention system (IPS) which will analyze the outgoing

and incoming traffic to the network. The firewall device needs to have this additional module to implement it. This feature analyzes the packet and filters the infected portion [12]. There are two kinds of firewall, and these are hardware based and software based. Software-based firewalls are generally installed in a host workstation and have so many drawbacks. Hardware-based firewalls are dedicated a device to operate security systems, and it can be set up in any position of the network. This ensures better protection and solution than software-based firewalls. At present, there are new kinds of firewalls, which is called as cloud-based firewall. These kinds of firewalls are easily accessed from remote and operated by specialized security professionals [9].

## 3 Methodology

Most of the organizations found out the existing security system controls are reducing their effectiveness and preventing them from getting something done. An organization needs to share some information with the general user, customer, or newfangled business partners. If there are any kinds of security issue occur, what will be the business impact? For the outcomes of this kind of security issue, it needs to be developed an innovative, effective security program. So, the risk depends on the security system and security policy that the administrator applied to the network. Sometimes good security of an organization creates a good business opportunity because the customer, business partners want to be sure that their data is secure as new risks are introduced daily, so the administrator needs to be aware of everything and should apply a proper firewall policy to prevent those attacks and risks. The company needs to balance the security issue and business information while sharing company information can share information easily [9] (Fig. 1).

### *3.1 IP Spoofing*

Ip spoofing is to obtain unauthorized entree to the computers which are a kind of technique that attackers use. In this process, the hackers send illegitimate messages to the host computer IP as a trusted host. On another hand to generate traffic in the target computer and make it coming from a trusted host attacker engages with the unsuspected hosts, then the attacker floods the network. The way of transmitting packets with forged source addresses is known as IP spoofing, which is straight linked to different kinds of networks malfunctioning like DDoS [13, 14]

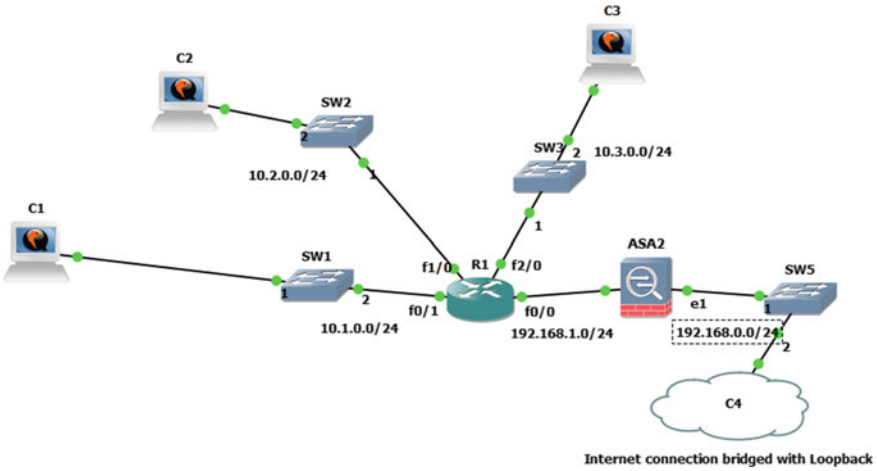



**Fig. 1** Existing network model of an enterprise network

## *3.2 Insider Intrusion*

Insider intrusion is kind of unusual type of attack; it is not like an external threat, a person who has authorized system access could be an insider intrusion attacker, so there may be less security against insider intrusion on most of the organization because most of the company focus on their external attacks, and insider attack is also recognized as insider threat [15].

## *3.3 Denial of Service (DDoS)*

The DDoS attack is distributed for a server, website, or another network resource by multiple computer systems attack. The target system forces it to delay or even crash and even shut down as a flood of an incoming message, connection request, or malformed packets to the target [16].

## *3.4 No Protection Against Masquerades*

A masquerade attack happens by using the fake identity of a network or system to acquire unofficial access to personal or organizational data. The attacker tries to steal user passwords and other confidential information. To prevent masquerade, attack some security features must be enabled for authorization and proper authentication.

### *3.5 Firewall Trusted and Untrusted Network (LAN & WAN)*

In a firewall, the LAN is by default accepted as a secure network, and WAN is insecure. For that reason, security architectures named the LAN as trust and the WAN as untrusted. The trust network is defined as inside network, and the security level is higher where the untrusted network is defined as outside, and the security level is lower. It is allowed to initiate traffic from a higher security level to lower. In another way, around traffic initiation from lower to higher works opposite [17]. Security measures are appropriate or not depend on the threat profile that helps us to understand network management for a self-connected LAN network; here, no need to have network management protocol encryption or special authentication for those protocols, and the network administrator does not want his network management protocols to traverse without the special authentication of Internet protocols; so for any system, first step is to identify the threat and then apply threat prevention policy to ensure security protection.

The proposed model may overcome the above all existing problems of an enterprise network. Here is those probably improved part of the model given below:

- Strong Authentication.
- High Data Security.
- Multilevel Protection.
- Network Traffic Encryption.
- No Insider Intrusion.
- Strong Masquerades.
- IP sec used.
- Port forwarding.
- Internet traffic filtering.
- Different web access policies for users.

## 4 Implementation

Here, the proposed research work is focused on the main things that are required to implement and enhance the security model of our proposed system of an enterprise network. The purpose of this security of an enterprise network is to protect its valuable data, file, document, and many other things from attackers, hackers, and other dangerous platforms. To prevent those attacks and enhance our security, a proper authentication and authorization are required. Section 1 will observe what kind of tools are needed for this process to implement in the enterprise networks and proper evaluation. Section 2 will overview the process of policy to apply like NAT, ACL, WAN, and LAN zone, trust, and untrusted network determine and various things. Section 3 will illustrate the port forwarding in the LAN side to secure our potential websites, server, different devices that are accessible by the Internet with a major concern of security. Section 4 will view how IPsec VPN policy is applied for higher

security purposes and better authentication from vulnerability, and packet spoofing is resolved from active and passive attacks. And at the end of this sector, bandwidth utilization of this proposed system that is required to implement an enterprise network is analyzed.

## *4.1 Devices and Appliances*

To enhance our proposed model, a network architecture of our desire system is drawn and makes a topology in graphical network simulator-3 (In Short GNS3). This part considers three cisco router and two Fortinet firewalls and some non-manageable switch connected with the Internet cloud separately. In the existing system, the ASA 8.42 firewall is used which is costly and expensive and also have some limitation. There are different types and better-secured firewalls in our internetworking world to secure the world of the Internet now in this era. Some of these firewalls like Cisco ASA, Fortinet, Fortinet-Hyper version FortiGate firewall, Sophos firewall, software firewall (Like ACL), Barracuda spam firewall, etc., are available. The FortiGate firewall will be implemented in this system to enhance our security. This firewall has excellent features with flexibility and comfortability both graphically and CLI mode. Fortinet with the advanced version of FortiGate has various update versions that are user-friendly and highly secured to protect our LAN and WAN portion on the Internet. Free meter tools are used for bandwidth monitoring and bandwidth utilization of our existing system as well as how much improved in our proposed system to obtain better results and help to find out the effective results and also inefficiency measuring. This tool installed in host pc that means any change in bandwidth gives us the result of bandwidth utilization. This free meter tools are designed for the Windows operating system which is almost reliable for bandwidth monitoring and also for the process of evaluation. WinMTR is one of the free tools to lookup traceroute of a networking device connected with the Internet. Having a connection with the proper establishment with actual route from source to destination can easily give us traceroute hop to hop and also ping statistics of a network. This is also called network diagnostics tools that are truly named after Matt's traceroute. A network connection with the proper route in topology or a diagram facing any network difficulties or slow speed can be determined by the WinMTR. Fortinet FortiGate hyper version is a next-generation firewall with advanced features and technology that is used for higher security in the LAN and WAN side as well as secure our network connection in the world of the Internet. In my opinion for an enterprise network, the Fortinet firewall is the best solution to fulfill or requirement and meet up the challenge of privacy and protection. The firewall leads the maximum efficiency in both platforms (Virtualized and Hardware).

**The devices and tools are used in this topology are given below**:

**Network Design**: The network or topology of our proposed system of an enterprise network is given below compared to an existing system that is already showed. In

**Table 1** List of tools and devices for implementation and application

| Device name | Specification | Installed tools |
|---|---|---|
| Host PC—PC1<br>PC2<br>PC3 | CPU: Core i4,<br>1.5 GHz | Securecrt 8.5.2,<br>WinMTR,<br>FreeMeter |
| Cisco switch | Layer 2 device,<br>Non-manageable | GNS3 |
| Router c7200p | Layer3 device,<br>Fast ethernet port,<br>Gigabit port,<br>Maximum speed accuracy | Cisco c7200 iOS,<br>Wireshark |
| ASA 8.42 firewalls | Identity firewall,<br>Identity NAT<br>configurable proxy<br>ARP and route lookup | Cisco ASA<br>8.42 iOS,<br>ASDM bin |
| Firewall | Fortinet -FortiGate<br>Hyper versions | FortiGate VM,<br>FortiGate-3140B |
| Metasploit<br>Framework | Penetrating tools | Windows 10<br>OS |

this topology, it is a scenario of an enterprise network where corporate office and branch office connected to head office in the same network. In this topology, the internet cloud is connected with a Fortinet firewall to ensure security, privacy, and authentication.

Then, this head office firewall is connected with another Fortinet firewall with encryption and using Internet protocol security (IPSec) VPN for higher security purposes along with the synchronization with routers, switches, and L3 switches consecutively with end device (host PC). This is a common scenario of our proposed network to enhance the security of an enterprise network (Table 1).

## *4.2 Implementation of Firewall with the Integration of Different Policy*

This section describes the different policies applied in Fortinet firewalls and various steps taken to remove the different attacks as those attacks mentioned before and how those attacks mitigate in FortiGate firewall. As in this network topology where the whole network is scanning, by the Wireshark software find the attacks before the action is needed from the firewall. For implementing security, both the results from the network with its scanning process are compared to show whether the reconnaissance. Figure 2 shows the FortiGate-3140B firewall graphical mode of IPS page with an observation of the Aurora attack is detected and needs to prevent it. Aurora attack

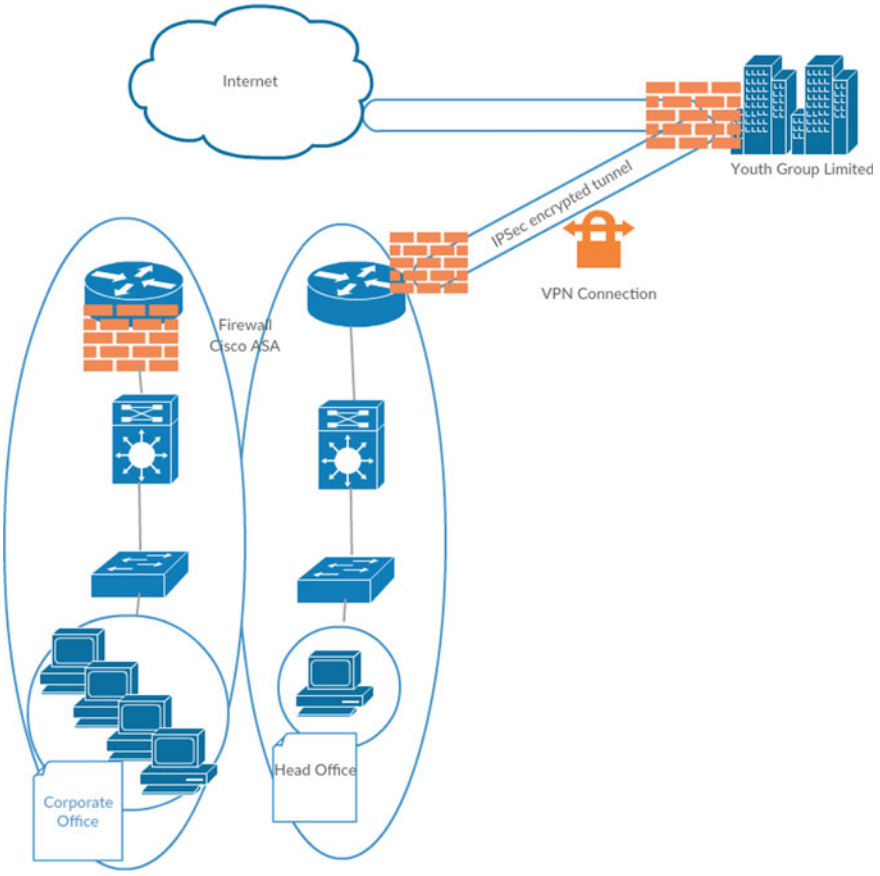



**Fig. 2** Proposed model for an enterprise network

is highly harmful to the entire system. It targets s on the server host that takes place in the Windows operating system as well as various random places that are essential for protecting data.

For removing attack and secure our data, it has been required to protect HTTP, HTTPs, Telnet, SSH, and SNMP port in the firewall as well as in the router. So, it is essential to configure and monitor the firewall carefully. Setting up the FortiGate firewall and configured it properly. Figures 2 and 3 show that the iOS file configures in the VMware and established a connection host PC and virtual PC connected and setting IP and port in the FortiGate firewall in CLI mode. When all the processes of configure IP and port are successfully configured, then the graphical mode of FortiGate firewall can be accessed. The assigned IP in the port of firewall is having a smooth ping statistics from source to destination.

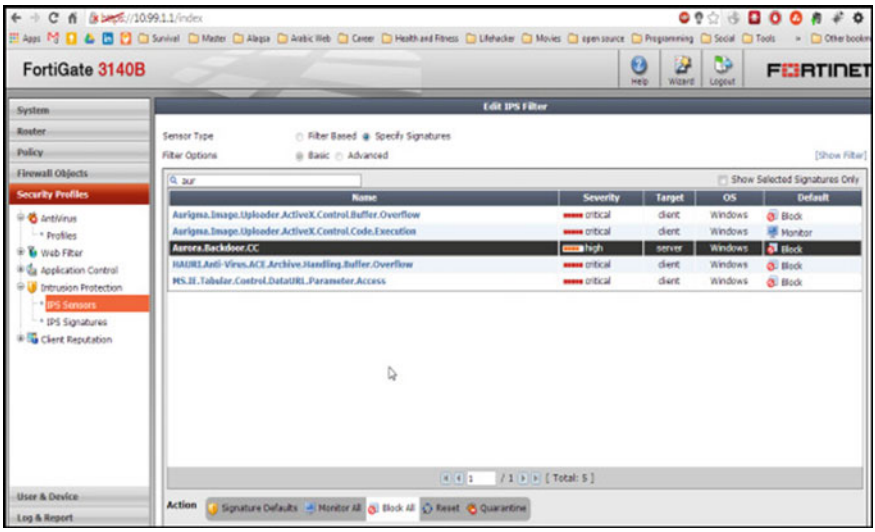



**Fig. 3** Aurora attack takes place IPS page FortiGate-3140B

## *4.3 Implementing Port Forwarding and Policy Against Direct Internet Traffic*

Port forwarding on router or firewall or any kind of device that controls or monitors traffic on the Internet allows a port address to enter into it. Port forwarding plays an important role to secure vulnerable port of potential devices form attackers and hackers to keep safe our important data. With port forwarding, the internal port of a router or firewall along with an IP address can be changed where port number forwarded to unknown port for incoming. This is a tricky way to secure and put privacy on the firewall to protect our data. This firewall, TCP port 21 is forwarded to 8000 with the private IP mapping with public IP, as it gave from Internet service provider. Port forwarding or port mapping is part of network address translation (NAT) where it is applicable. Direct Internet Traffic: Direct Internet traffic is that a hacker sends malicious virus and different kinds of direct attack through third party software like uTorrent, Bit Torrent, and much other open-source software in our window operating system with the permission of access firewall without unknowing of mind.

This is an important problem and a security concern nowadays. It can be removed by our awareness when clicking any harmful link or installing any crack open-source software in our Windows operating system. So, the rules and policies like objects (IP/Subnet, FQDN) create a FortiGate firewall and introduced them in a particular declaration with these specific objects. These objects are blocked from the firewall were trusted and untrusted zone of LAN sides. Some of the objects created FortiGate that is objects are restricted from firewalls when accessing some untrusted sites any

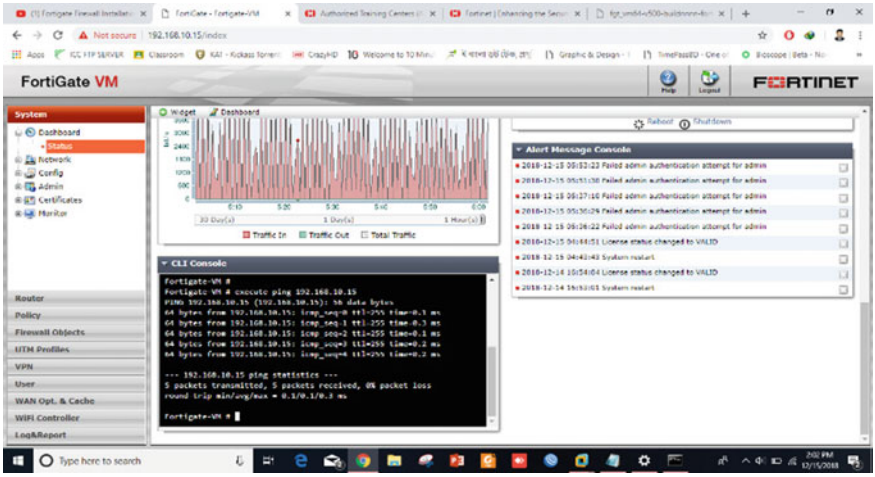



**Fig. 4** Ping statistics of FortiGate VM with host PC

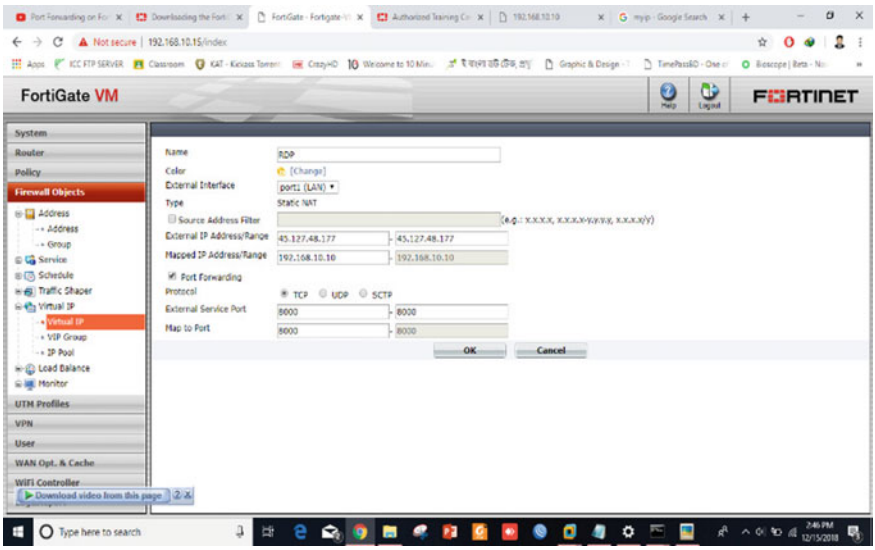



**Fig. 5** Port forwarding in FortiGate firewall

suspicious websites. Then, the policy applied for direct Internet traffic to remove malicious attacks from different ports, open-source software, and create a strong zone for protecting valuable data from attackers. And the traffic for direct Internet monitoring shows in Figs. 4 and 5, in that monitoring gives us a concept or knowledge about direct traffic filtering in incoming and outgoing. From the figure, it can be understood that the problem occurs in direct Internet traffic.

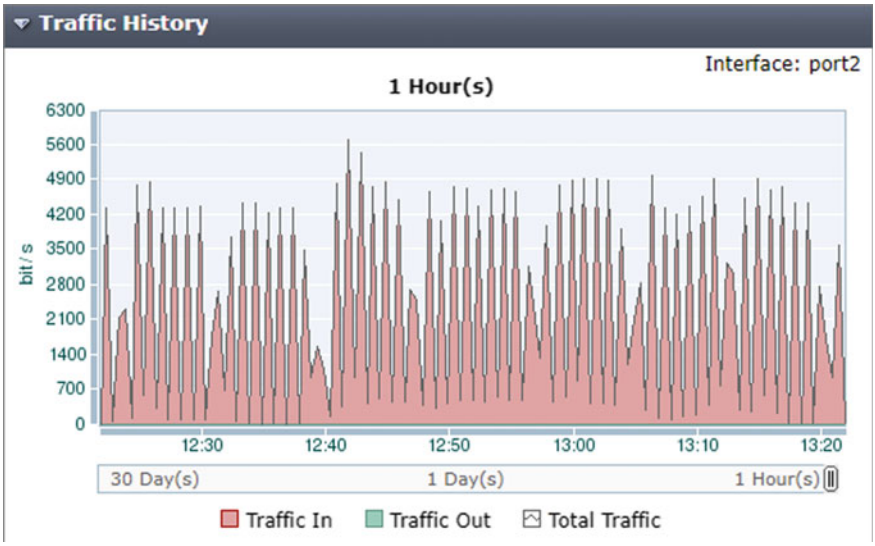



**Fig. 6** Traffic history of direct Internet traffic ingoing and outgoing

## *4.4 Implementation of Internet Protocol Security (IPsec) VPN in Proposed Model*

The IPsec is used for better security and IETF standard protocol. It maintained the encryption process in sending data from one place to another. The IP layer is being secured when data is sending with an encryption process. A cryptographic key and also pre-shared keys use during the session. The proposed modes have applied IPsec VPN between two FortiGate firewalls, which have a secure process in a big enterprise network to prevent intrusion and different kinds of threats. For a session with end to end security, IPsec VPN provides the best TCP/IP layer security along with a higher level of authentication. In WAN port LAN port where there is a route between source and destination, a pre-shared key is used for remote connection in FortiGate firewalls to another destination firewall with a virtual private connection is established. A tunnel is established between head office and the corporate office with IPsec VPN to secure data that are passing through internet. The tunneling creates a strong zone in the internal network having several conditions of firewall policy with file sharing or data coming from the outside network as well.

Another IPSec VPN configures in the remote destination in the same way along with pre-share keys encryption, a remote destination address, and port (LAN) selection. Figures 6 and 7 show the configuration of IPSec VPN for destination host. The best path chooses to give encryption to make a season with local host. And the firewall policy is applied in the FortiGate firewall.

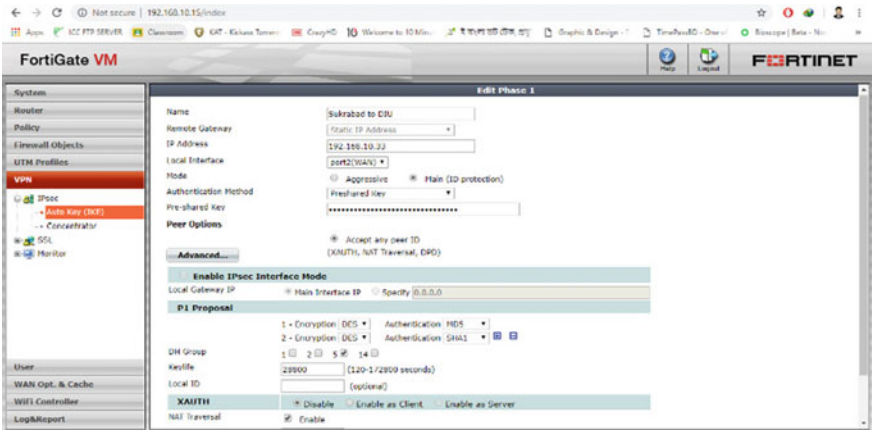



**Fig. 7** VPN configuration in local interface

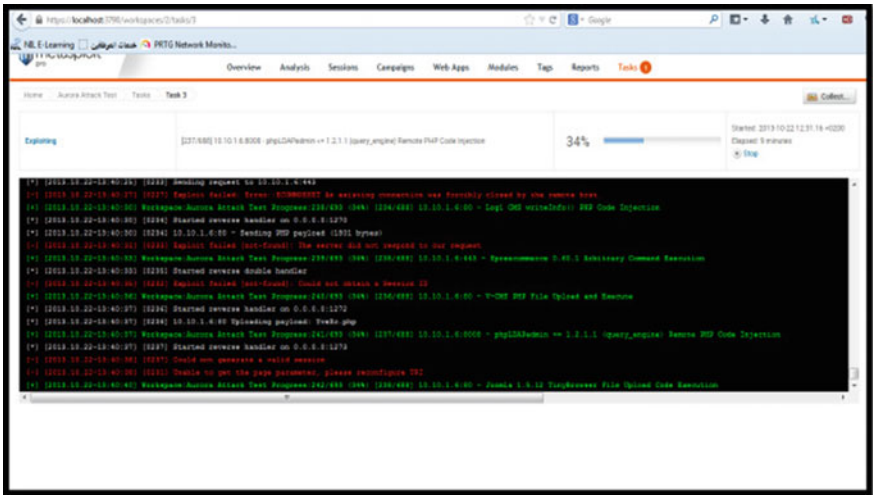



**Fig. 8** Lunching Aurora in metasploit tools with port forwarding

## 5 Results

This section will evaluate the results of applying the different policies, rules, and various configurations in the Fortinet FortiGate firewall. The following figures show the result of port forwarding, protection against DoS attack, direct Internet traffic, port scanning, and IPsec VPN. This evaluation is for the proposed system of an enterprise network. Using the Metasploit tool for the penetrating attack in the FortiGate firewall. And this penetrating attack is done with various purposes like port scanning and Aurora attack for evaluation of log reports in the FortiGate firewall. Figure 8 shows

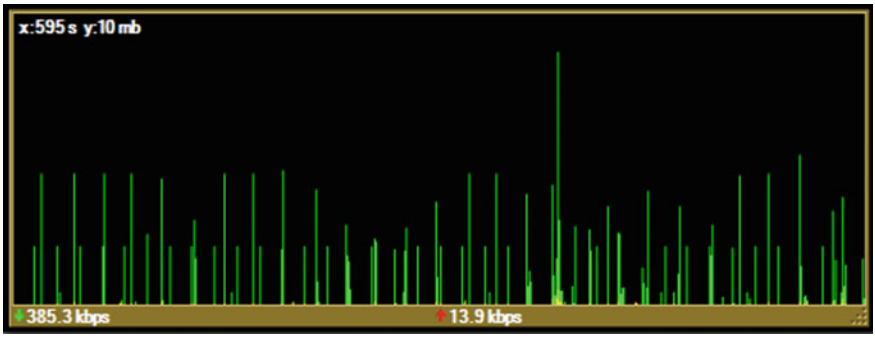



**Fig. 9** Utilization of bandwidth before applying the proposed model

penetrating attack by the metasploit tool to detect Aurora attack, port scanning in the FortiGate firewall. In our paper we have checked the vulnerability in firewall by using internet explorer with local hosting.

Some object of Aurora attack with IP address and port in the FortiGate firewall. The vulnerability scan and log report generated from FortiGate firewall is given in the Fig. 8.

## *5.1 Performance Analysis and Evaluation*

The free meter tools are used for monitoring, and the utilization of bandwidth in the proposed system of an enterprise network is compared to the existing system. One more thing to say, Fortinet firewall is comparatively cost-effective and reliable than adaptive security appliances (ASA). On the other hand, ASA firewall is costly though reliable when it has been taken into consideration to design a network topology for an enterprise network; firstly, it will consider how much it is cost-reducing when buying the network equipment. An effective network topology not only helpful and necessary for enterprise networks, but also efficient and needful for a company. When different policies and rules are applied in the firewall to protect our data from attackers or hackers, then bandwidth is also a matter of concern. The free meter tools are used for measuring and mentoring bandwidth utilization when all the processes running in the firewall. Figure 9 shows bandwidth utilization before applying the proposed model, and Fig. 10 shows the bandwidth after using firewall as the procedure already mentioned before in this paper.

So comparatively, bandwidth utilization in our proposed monitoring system is better than the existing system considering maximum usages of maximum traffic. And the obtained CPU usages is moderate than the existing system. If the attacks are differentiated between the proposed system and existing system in a chart of

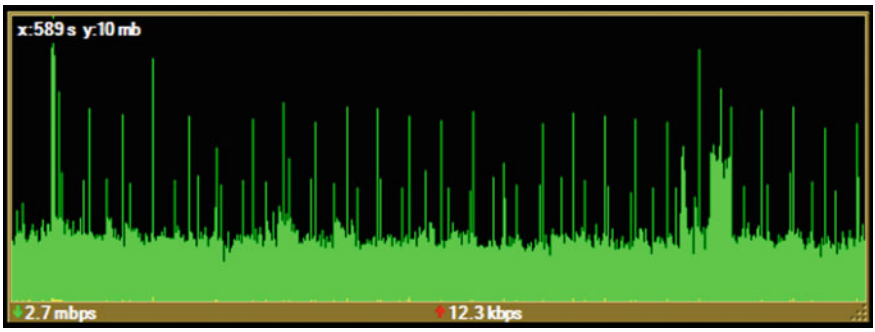



**Fig. 10** Utilization of bandwidth after using firewall

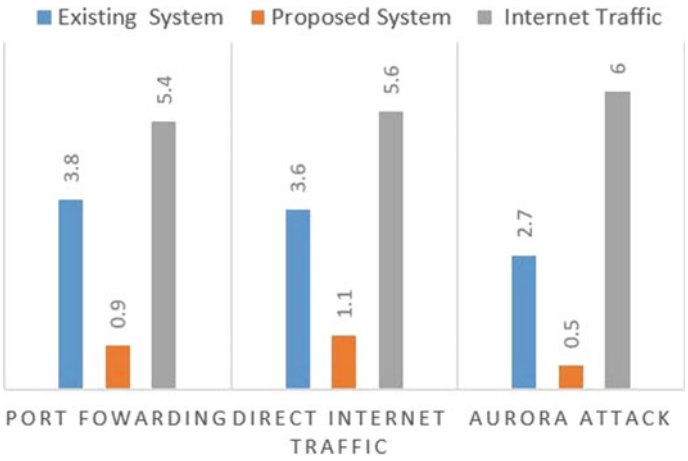



**Fig. 11** Comparison between existing and proposed system

graphical view, it will be easier to understand, and the anomaly of these system can be detected.

## 6 Conclusions

This paper has developed a network enhanced model for an enterprise network. Basically, an enterprise network includes a different device, network, equipment, and protocol. Different devices are connected in a different layer. In layer, two different devices like switches and printers are connected to the network. And layer three core switches like routers are connected to the network. These core switches are used to connect different branches in different networks. An enterprise network is attacked by many threats like external and internal threats. Also, attacks came from different

layers. To prevent this attack, enterprise network uses different devices like VPN and firewall. Security threats become a big challenge to create an enterprise network. The proposed network model in our paper is also an enterprise network. This enterprise network is affected by different threats like IP spoofing, phishing attack, DoS attack, spyware attack, etc. The proposed model gives a better solution to prevent the attack. As an enterprise network security issue is critical, firewalls and VPN are deployed to protect the network. This research work has deeply analyzed the proposed model to prevent different kinds of attacks. The results of our proposed security model prove that it has the ability to detect and protect the network from different kinds of attacks. But there is no guarantee that the proposed security model can detect the new attack. The proposed security model only detects some attacks to secure our enterprise network.

## References

1. Chen LC, Lin C (2007) Combining theory with practice in information security education. In: Proceedings of the 11th Colloquium for information systems security education, pp 28–35
2. Trabelsi Z, Ibrahim W (2013) A hands-on approach for teaching denial of service attacks: a case study. J Inf Technol Educ Innovations Pract 12:299–319
3. Bouhoula A, Trabelsi Z, Barka E, Benelbahri MA (2008) Firewall filtering rules analysis for anomalies detection. Int J Secur Netw 3(3):161–172
4. Trabelsi Z, Zeidan S (2012, June) Multilevel early packet filtering technique based on traffic statistics and splay trees for firewall performance improvement. In: 2012 IEEE international conference on communications (ICC). IEEE, pp 1074–1078
5. Trabelsi Z, Zhang L, Zeidan S (2014) Dynamic rule and rule-field optimisation for improving firewall performance and security. IET Inf Secur 8(4):250–257
6. Trabelsi Z, Shuaib K (2008) A novel man-in-the-middle intrusion detection scheme for switched LANs. Int J Comput Appl 30(3):234–243
7. Alnagi KW (2014) Developing security-enhanced model for enterprise network. Developing security-enhanced model for enterprise network
8. Gaigole MS, Kalyankar MA (2015) The study of network security with its penetrating attacks and possible security mechanisms. Int J Comput Sci Mob Comput 45(5):728–735
9. Ritchot B (2013) An enterprise security program and architecture to support business drivers. Technol Innov Manag Rev 3(8)
10. Taluja MS, Dua RL (2012) Survey on network security, threats & firewalls. Int J Adv Res Comput Eng Technol (IJARCET) 1(7)
11. FortiGate: Next-Generation Firewall (NGFW) (November 12, 2019, 9.24 PM) is available https://www.fortinet.com/products/nextgeneration-firewall.html
12. Adebayo SA (2012) Network security [Unpublished Bachelor's Thesis]. Turku University of Applied Sciences, Turkey
13. Singh J, Kaur L, Gupta S (2012) A cross-layer based intrusion detection technique for wireless networks. Int Arab J Inf Technol 9(3):201–207
14. Wang H, Jin C, Shin KG (2007) Defense against spoofed IP traffic using hop-count filtering. IEEE/ACM Trans Netw 15(1):40–53
15. Chagarlamudi M. Panda B, Hu Y (2009, April) Insider threat in database systems: preventing malicious users' activities in databases. In: 2009 sixth international conference on information technology: new generations. IEEE, pp 1616–1620

16. Shannon C, Moore D, Claffy KC (2002) Beyond folklore: observations on fragmented traffic. IEEE/ACM Trans Netw 10(6):709–720
17. Le Boudec JY (1992) The asynchronous transfer mode: a tutorial. Comput Netw ISDN Syst 24(4):279–309

# Comparative Study of Fault-Diagnosis Models Based on QoS Metrics in SDN


**Anil Singh Parihar and Nandana Tiwari**

**Abstract** Due to the current exponential rise in users of the Internet, a need for highly scalable and easily configurable network devices is required. To meet this growth in demand, software-defined networking (SDN) has become increasingly popular. However, despite their several advantages, there still exists a gap in the level of quality of service (QoS) required in the existing fault detection and recovery mechanisms to promote a wide-scale carrier-grade network (CGN) selection. In this study, the basics of SDN were reviewed and compared the five diverse failure recovery models as a function of their QoS.

**Keywords** Networking · Software-defined network · Quality of service · Carrier-grade networks · Network architecture · Fault detection · Restoration

## 1 Introduction

In the TCP/IP network architecture, routers and smart switches operate at the network layer. As smart switches work in a similar way to routers at the network layer, they will be referred to as routers as well from here on. Routers at the network layer consist of two main functions, namely routing and forwarding.

Routing refers to the computation done to ascertain the route a packet must take to reach from the source device to the destination device. As this function deals with the control of the routing device to route traffic through a particular path and mainly interacts in logic and software, it is known as the control plane.

On the other hand, forwarding refers to the actual displacement of a packet from the router's input to an appropriate output, based on the routing path computed earlier. Since this function deals with real data due to the packets and interacting with them on a physical or hardware level, it is known as the data plane.

Traditionally, both of these planes are an essential function of the devices operating at the network layer [14]. However, due to the importance of these functions and

A. S. Parihar (✉) · N. Tiwari
Delhi Technological University, Delhi, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_60

their interdependence on each other to execute the task of the network layer, much responsibility is allotted to these devices which can prove to be detrimental in many scenarios, the most obvious of which include failure in the software of these devices frustrating the role of the hardware and vice versa. Any change in the software of the device could also warrant a complete reconfiguration of it. Several other problems that may arise include, but are not limited to, the compatibility and extension of network protocols, switch software updates, network device maintenance and network innovations [1].

Hence, to overcome these failures, the need for devices operating at level 3 with a decoupled data and control layer arises. Software-defined networks (SDNs) support this by separating the control plane from the routers and hence reducing them to the function of forwarding packets alone. A centralized controller is installed globally, which performs the role of the control plane. This global controller provides support in many applications where the traditional network fails. For example, having a global control and view of the network, SDNs can easily find and configure an optimal alternate path in case the primary path of routing from a source device to a destination device was to face link failure. In contrast to this, the traditional network devices would flood the network with packets to compute an alternate route. It must also be noted that SDN defines a global controller that can use the overall view of the network to easily make optimal decisions; switches maintain connectivity amongst each other in the same way as in a traditional network [13]. A more detailed analysis on the components that make up the SDN are discussed in [11].

## 2 Background

### *2.1 Architecture*

SDN separates the data plane from the control plane. The three different layers are observed for concerning the transfer of data. At the lowest level of abstraction the data layer does the actual transfer of packets via devices such as switches. Above this layer is the control layer where the intelligent decision-making of routing takes place using the global view of the controller. Finally, the application layer is at the highest level of abstraction of the network, where all the application frameworks run.

Figure 1 depicts that there are two interfaces under consideration to facilitate the talking of one layer to the other [9]. The interface between the lowest layer of abstraction and the middle layer is called the southbound interface. Similarly, the interface between the highest level of abstraction and the middle layer is called the northbound interface. OpenFlow protocol is one of the most widely used southbound application programming interfaces (APIs) to allow the data and control plane to talk to each other as it enables openness and visibility in the network. The northbound interface also uses various types of APIs to allow the controller to talk to the applications running on top of it. These applications are often used to acquire global data of the

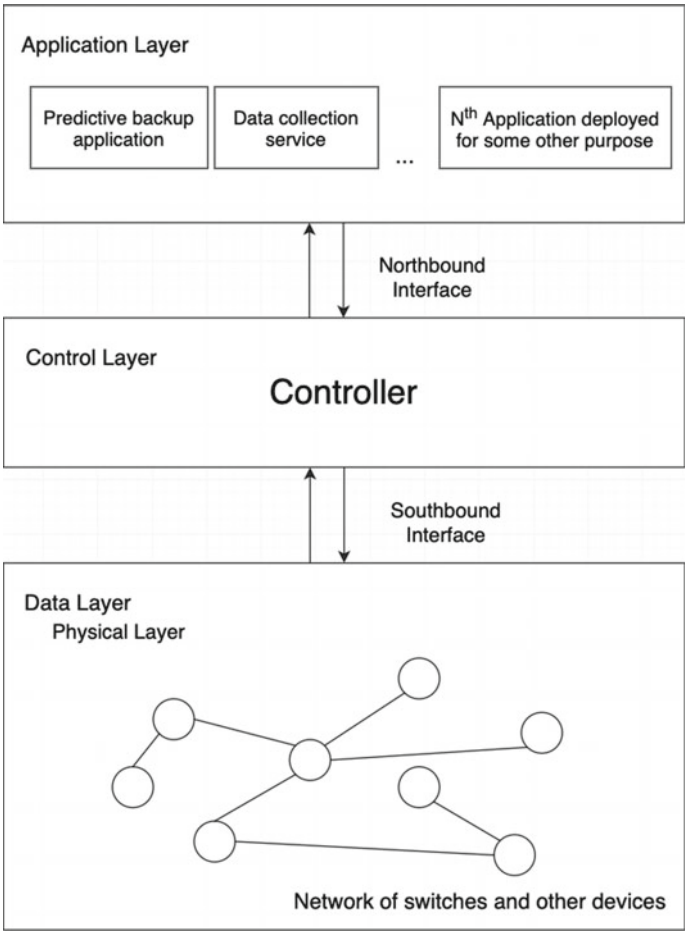



**Fig. 1** Architecture of SDN

network from the controller. These APIs include, but are not limited to, REST API, Onix API, Java API.

In an ideal scenario, switches will communicate with the controller using the southbound interface to send traffic information such as packet loss ratio and transmission time. Based on the applications running on top of it, the controller sends appropriate data to them through the northbound interface. For example, a predictive learning application may request the controller for traffic metrics to predict the most optimal path for a particular flow and send this information back to the controller through the northbound interface. The controller may then configure this most optimal path in the flow tables of the affected switches using the southbound interface.

### 2.2 OpenFlow

The OpenFlow switch consists of three main parts: flow tables, OpenFlow channel and the group tables [2]. The OpenFlow channel is used to talk with the controller(s) involved in managing the switch using the OpenFlow switch protocol. Each flow table consists of some flow entries that have three attributes: match field, counters, and a set of instructions to apply on matching. An OpenFlow switch may consist of multiple flow tables that are connected through a pipeline that allows the sharing of metadata across the flow tables [15]. The group tables are used to perform a more complex set of actions to the incoming packet, which are contained in action buckets.

When a packet reaches a switch to be forwarded to a destination, the switch searches its flow tables for a matching entry and follows the set of instructions defined in case of a match. It must be noted that the rules in the flow tables are defined according to priority, so if a matching entry is found, the switch will not look for other potential matchings. The set of instructions to be acted on the packet can range from being a forwarding rule to performing a lookup in the group table. In the case that no such matching exists, the packet may be forwarded to the controller through the OpenFlow channel. The controller then adds new rules in the form of flow entries to define the path the packet should be routed on. The study [10] presents an Open Framework for OpenFlow Switch Evaluation (OFLOPS) that permits the development of tests for OpenFlow-enabled switches and measure the capabilities and bottlenecks between the forwarding engine of the switch and the remote control application.

### 2.3 Fault Diagnosis

The main types of faults that may disrupt the service of the network are defined below [3]:

(a) Link Failure
This refers to the failure in the links connecting hosts to switches, switches to switches, switches to the controller(s) and controllers to controllers (if applicable).
(b) Controller Failure.
This refers to the failure of a controller that is managing the network.
(c) Device Failure.
This refers to the failure of a router that is responsible for forwarding traffic.

As discussed in the study [4], the frequency of device failure is less than the frequency of link failure. Hence, in this study, the first two types of failures were compared. In the case of a link failure, the recovery of a failed link can be achieved in two possible ways: reactive restoration method and proactive restoration method.

# 3 Key Concepts and Terminologies

## *3.1 Reactive Restoration Method*

In this section, the first approach called reactive recovery was explored. In this approach, on the occurrence of a link failure, the controller is approached for finding an alternative path for the flow that was affected and installing it in the affected switches. In OpenFlow, this is achieved by inserting new flow rules in the flow table of the OpenFlow switches. As the controller is contacted after the failure of a link, this approach may lead to delays. Apart from this communication overhead delay, another major contributor to the delay in restoration is the time required for the controller to compute the new optimal path. Due to this delay, this approach does not support speedy link fault recoveries; however, due to its simplicity, this approach can be scaled easily.

## *3.2 Proactive Restoration Method*

This section describes the second approach to link fault recovery. In this method, all the possible alternative paths are computed and configured in the switches as backup paths and can be used in case of link failure without any delay. Although this approach provides faster recovery of flows, it is hardly scalable. As the size of the network increases, there would be an exponential rise in complexity to configure each possible path as an alternative path. This exponential rise in complexity would be observed not just in computation but also in the space that would be required to store the flow entries. Hence, it can be concluded that the reactive approaches to be more time complex and proactive approaches to be more space complex; i.e., a tradeoff is observed.

## *3.3 Carrier-Grade Networks*

The most obvious way in which CGNs differ from other networks is that it faces a higher number of users due to their unwavering quality of service support for various protocols (e.g. VOIP applications, FTP applications). Fast failure recovery guarantee within a fixed time interval is essential for providing a service guarantee. Due to this, the CGN is governed by exceptionally high availability and reliability ($\geqq$99.999%). For an SDN system to be used as CGN, it must guarantee an unscheduled fault recovery within 50 ms. The study [8] demonstrated that a reactive restoration approach in OpenFlow could not be achieved within the Carrier-Grade Network level, whereas a proactive restoration approach could.

**Table 1** Quality of service metrics

| QoS metric | Definition |
|---|---|
| Throughput | It is a measure of the amount of data that can be successfully received within some unit of time |
| Bandwidth | It is a measure of the maximum amount of data that can be transferred within some unit of time |
| Availability | It is a measure of the total time the network is operational against the full time it is observed, i.e. the total uptime of the network |
| Tolerance | It is a measure of how operational the network as a whole is on the occurrence of some failure |
| Resilience | It is a measure of the level of performance the network can maintain on the occurrence of some failure |
| Scalability | It is a measure of how well a network can perform to a rise in the number of resources being used by it |

### *3.4 Quality of Service Metrics*

Although SDNs seem to satisfy the needs of modern-day networks, much research is continuing to enable them to meet CGN level of QoS with regard to failures in the network. QoS refers to the measure of how well a network performs when posed with constrained network capacity. Some of the metrics that can be used to ascertain the QoS of a network are defined in Table 1.

## 4 Comparison

In this section, the five diverse models were compared against each other. A comparative QoS analysis of these models is done in Table 2.

### *4.1 Fast and Adaptive Failure Recovery Using Machine Learning in Software-Defined Networks (FAFRs) [5]*

This model aims to use traffic metrics to configure backup paths for a particular flow adaptively. It employs a proactive approach to compute the goodness of a path based on the environment of the network in a specific instant of time. A flowchart of its working is shown in Fig. 2. This model also compared various classifiers against each other for prediction of the goodness of a path such as decision trees, random forest, linear regression, support vector machines, neural network and found decision trees and random forests to outperform the other classifiers significantly. Since this approach aims to configure the best backup path, which is disjoint from the primary

**Table 2** Comparison of the five models based on their quality of service

| | FAFR [5] | PRT [6] | FT-SDN [3] | CN [7] | PFR [2] |
|---|---|---|---|---|---|
| Controller used | Floodlight | POX | OpenDaylight, floodlight, Ryu | NOX | POX |
| Recovery plane | Data plane | Data plane | Control plane | Data plane | Data plane |
| QoS optimized | Bandwidth | Availability | Resilience in controller failure; load balancing of controllers | Scalability; faster than basic reactive recovery | Packet loss; faster than basic proactive recovery |
| QoS degraded | Recovery time; scalability | İnstability | Packet loss | Inconsistency with respect to support to applications | Bandwidth scalability |

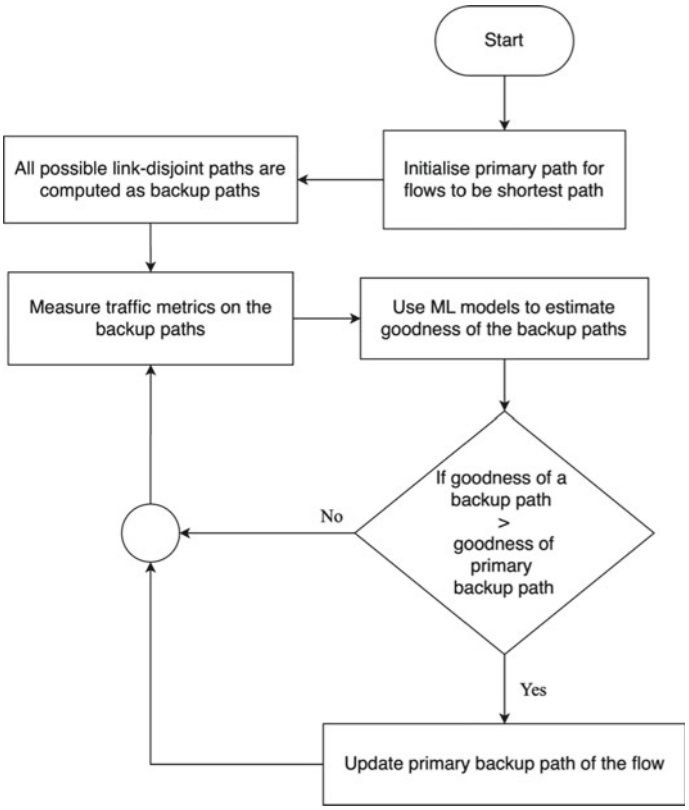



**Fig. 2** Logical flow of events of FAFR

path based on traffic metrics, it also supports higher bandwidth allocation per flow when compared with a baseline approach. This model reduced failure recovery time by 50 and increased its network bandwidth allocation by 24%. However, as the number of nodes increases, the complexity of this model increases as well. This is because the goodness of each disjoint path for a flow is calculated to find the most optimal backup path. Such a procedure could result in a considerable overhead with the growth of the network. Hence, this model does not scale well. Also, since this model aims to select the path with the most favourable traffic metrics as an optimal backup path, this could result in choosing a path that may not be the shortest path between a source–destination pair. Hence, this approach has a higher number of hops for each flow in comparison with a baseline model, by an approximate of 2 hops.

### *4.2 A Proactive Restoration Technique for SDNs (PRT) [6]*

This paper uses a model similar to the one described in the study [12]. The model proposed here is also proactive, and its main aim is to make the network highly available. It accomplishes this by predicting the failure events before their occurrence by classifying paths as risky if their probability of failure crosses a threshold value. This path is then replaced by the shortest disjoint path. In case the prediction is correct, the flow corresponding to the replaced path is inserted into a failed flow array for which Dijkstra's algorithm is run to compute the shortest route for the flow after the failure. The flow of events for this model is described in Fig. 3. However, in case the prediction is wrong, the network rolls back to its previous state. Hence, this article evaluates the model's performance based on availability and instability of the network, where the availability is prioritized at the cost of network instability. Predicting the occurrence of a fault affords the network a specific time segment to prepare for the fault to occur. As a result, the measured service availability of their network is 97%. However, the frequency of the link failure events are generated and injected into the system based on a probability function; i.e., they are not simulated against real fault events. Furthermore, performance metrics evaluated by this model are routing flaps (a measure of the instability of the network) and availability alone; i.e., metrics such as time delay for the routings are not considered.

### *4.3 CORONET: Fault Tolerance for Software-Defined Networks (CN) [3]*

Here, the main aim is to optimize data plane fault tolerance. It does this using a reactive approach which includes four main modules. These modules are responsible for discovering topology, calculating backup path in case of a link failure, configuring the routing path, and assigning host traffic to the routing path, respectively. It

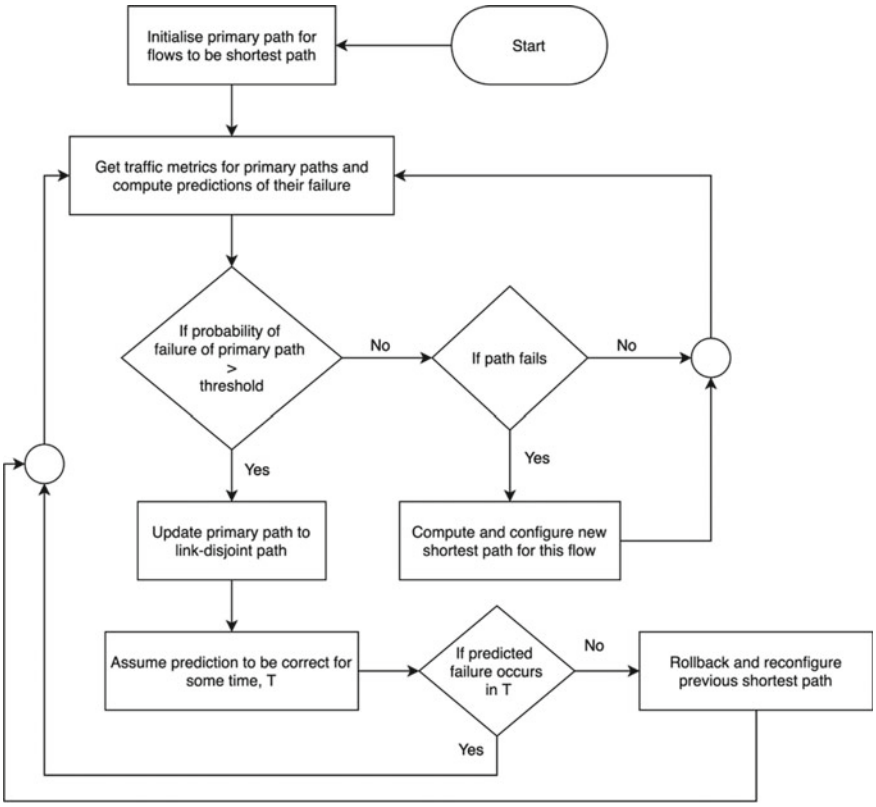



**Fig. 3** Logical flow of events of PRT

avoids controller intervention altogether in the instances of discovering topology and detecting link faults; these responsibilities are handled by the switches based on local switch standards such as Link Layer Discovery Protocol (LLDP). Once the packets arrive at an edge switch, it uses VLAN to forward them to their destination. The flow of events for this model is shown in Fig. 4. As there is no controller intervention, this is a fast recovery method. It is also highly scalable since local switches handle much of the time complex activities. However, experimental evaluation concerning any performance metrics was not shown, and the use of logical paths to route the traffic may prove to be inconsistent in some applications that already use physical paths for routing.

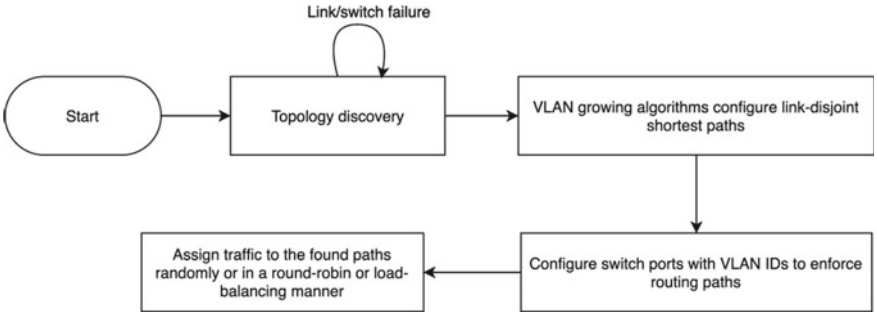



**Fig. 4** Logical flow of events of CN

### *4.4 FT-SDN: A Fault-Tolerant Distributed Architecture for Software-Defined Network (FT-SDN) [7]*

This is the only controller failure model that was analysed in this study. This model deals specifically with failures in the control plane. It aims to bring about network control plane resilience in the case of a failure. This is achieved by establishing an absolute distributed configuration of heterogeneous controllers. Each switch in the network is allotted a controller based on the distance and cost (in terms of load) between the switch and the controller. The controller selected is called the serving controller, and the rest are called secondary controllers for that switch. A database and metadata are maintained at each controller's end where it stores the states of the switches it is connected to. In the case of a controller failure, a new controller is selected to take over the failed controller's switches. The new controller is given the database information of the failed controller. The mechanism that enables this transfer of data from one controller to the other is done through the state replication method (SMR). The architecture proposed in this model also mandates a necessary amount of redundancy of state using the SMR; each serving controller updates its state to the other secondary controllers using the SMR. Figure 5 shows a flowchart explaining the flow of events in this architecture. As this is a truly distributed architecture of controllers, it is hence resilient to control plane failures. The use of heterogeneous controllers also increases the capability of the network to handle language-specific failures. However, the model deals only with the faults in the control plane and does not take link faults in the data plane into consideration. Also, if a controller fails and all other controllers are highly loaded, the model described in this paper is programmed not to overload the other controllers. Hence, packets from the switches that the failed controller was handling get dropped, lowering the resilience of the data plane of the network.

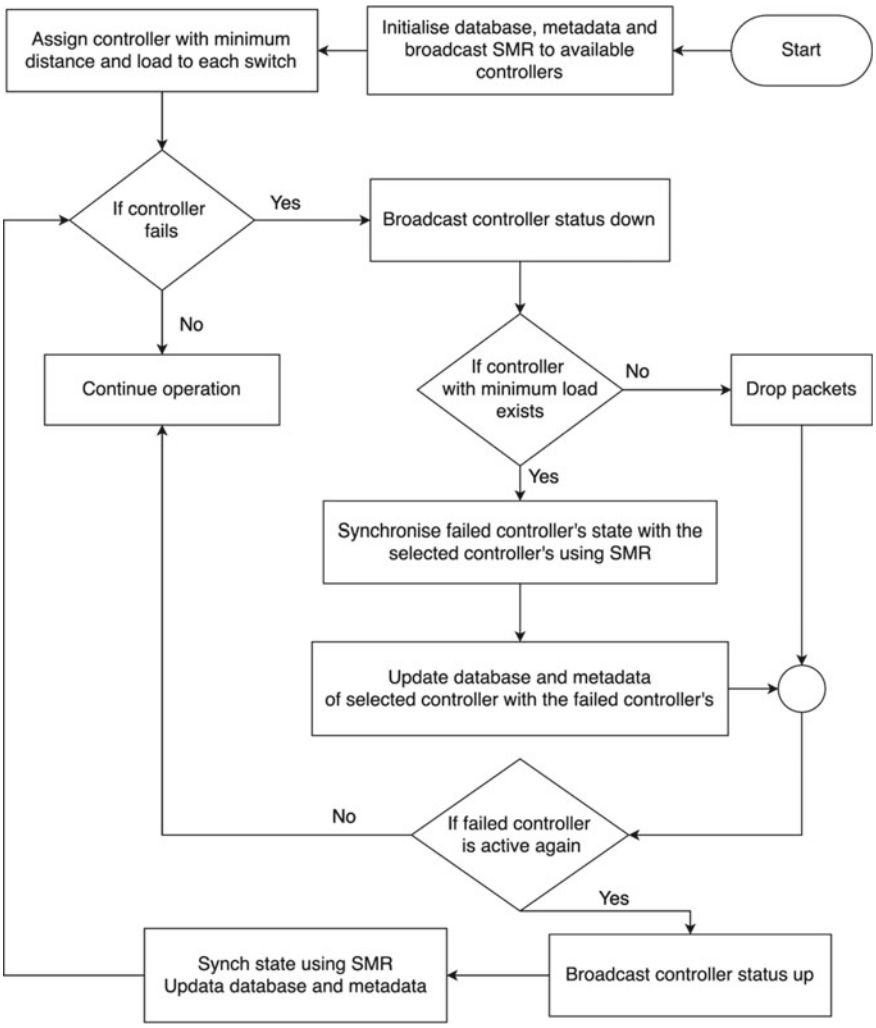



**Fig. 5** Logical flow of events of FT-SDN

## *4.5 Proactive Failure Recovery in OpenFlow Based Software Defined*

### 4.5.1 Networks (PFR)

The model proposed in [2] aims to provide link protection over path protection using four significant modules, similar to the previous model [3]. Path protection is defined here as the approach in which primary and backup paths are precomputed and installed in the flow table. Link protection is defined as the approach to precompute

and instal backup paths for each link. In this model, the only module that exists in the controller is the recovery module, which is responsible for computing backup paths of failed links. The rest of the modules are based on the data plane. They are responsible for detecting link failure, generating packets for allowing flow rules to persist, and for sending restoration packets to the controller to enable it to compute the shortest routing path with the new backup link operating as primary. This model provides an optimally fast recovery time of link failures and a negligible packet loss ratio, much better than the packet loss ratio over a path protection scheme. However, due to the multiple backup paths being installed in the flow tables, lesser bandwidth would be utilized by the network. It is also not a scalable solution as each link has multiple backup paths installed which would flood the flow tables with an increase in network nodes (Fig. 6).

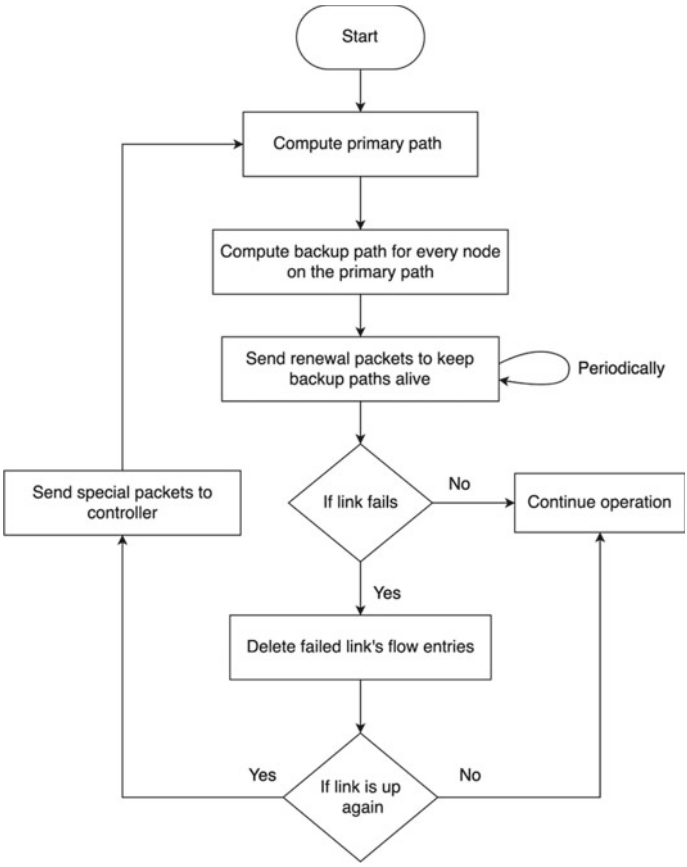



**Fig. 6** Logical flow of events of PFR

## 5 Conclusion

In this study, the need for software-defined networks to meet CGN levels of QoS was explored. The major domains of faults that may occur in an SDN and the proactive and reactive approaches that can be used to combat these failures were also explored. Finally, five diverse models were compared against each other for the QoS metrics, optimization and degradation, since a tradeoff would always be maintained. Based on this analysis, different models can be employed to accommodate different use cases.

## References

1. Yu Y, Li X, Leng X, Song L, Bu K, Chen Y, Yang J, Zhang L, Cheng K, Xiao X (2018) Fault management in software-defined networking: a survey. IEEE Commun Surv Tutorials. 1. https://doi.org/10.1109/comst.2018.2868922
2. Official OpenFlow Documentation (version 1.3.3) https://opennetworking.org/wp-content/uploads/2014/10/openflow-spec-v1.3.3.pdf
3. Kim H, Schlansker M, Santos JR, Tourrilhes J, Turner Y, Feamster N (2012) CORONET: fault tolerance for software defined networks. In: 2012 20th IEEE international conference on network protocols (ICNP), Austin, TX, pp 1–2. https://doi.org/10.1109/icnp.2012.6459938
4. Gill P, Jain N, Nagappan N (2011) Understanding network failures in data centers: measurement, analysis, and implications. In: Proceedings of the ACM SIGCOMM 2011 conference (SIGCOMM '11). Association for computing machinery, New York, NY, USA, 350–361. https://doi.org/10.1145/2018436.2018477
5. Truong-Huu T, Prathap P, Mohan PM, Gurusamy M (2019) Fast and adaptive failure recovery using machine learning in software defined networks. In: 2019 IEEE international conference on communications workshops (ICC workshops), Shanghai, China, 2019, pp 1–6. https://doi.org/10.1109/iccw.2019.8757169
6. Malik A, de Fréin R (2020) A proactive-restoration technique for SDNs. In: 2020 IEEE symposium on computers and communications (ISCC), Rennes, France, 2020, pp 1–6. https://doi.org/10.1109/iscc50000.2020.9219598
7. Das RK, Pohrmen FH, Maji AK et al (2020) FT-SDN: a fault-tolerant distributed architecture for software defined network. Wireless Pers Commun 114:1045–1066. https://doi.org/10.1007/s11277-020-07407-x
8. Sharma S, Staessens D, Colle D, Pickavet M, Demeester P (2013) OpenFlow: meeting carrier-grade recovery requirements. Comput Commun 36:656–665. https://doi.org/10.1016/j.comcom.2012.09.011
9. Kreutz D, Ramos FMV, Veríssimo PE, Rothenberg CE, Azodolmolky S, Uhlig S (2015) Software-defined networking: a comprehensive survey. Proc IEEE 103(1):14–76. https://doi.org/10.1109/JPROC.2014.2371999
10. Rotsos C, Sarrar N, Uhlig S, Sherwood R, Moore AW (2012) OFLOPS: an open framework for OpenFlow switch evaluation. In: Taft N, Ricciato F (eds) Passive and active measurement. PAM 2012. Lecture notes in computer science, vol 7192. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28537-0_9
11. Papagiannaki K, Argyraki K, Ballani H, Bustamante F, Camp J, Chaintreau A, Gill P, Mellia M, Raman B, Sommers J, Viana AC (eds) (2015) SIGCOMM'14. SIGCOMM Comput Commun Rev 44(4)

12. Vidalenc B, Ciavaglia L, Noirie L, Renault E (2013) Dynamic risk-aware routing for OSPF networks. In: 2013 IFIP/IEEE international symposium on integrated network management (IM 2013), Ghent, 2013, pp 226–234
13. Ali J, Lee GM, Roh BH, Ryu DK, Park G (2020) Software-defined networking approaches for link failure recovery: a survey. Sustain MDPI Open Access J 12(10):1–28
14. Yu Y et al (2019) Fault management in software-defined networking: a survey. In: IEEE Commun Surv Tutorials 21(1):349–392. https://doi.org/10.1109/comst.2018.2868922
15. Padma V, Yogesh P (2015) Proactive failure recovery in OpenFlow based software defined networks. In: 2015 3rd International conference on signal processing, communication and networking (ICSCN), Chennai, pp 1–6. https://doi.org/10.1109/icscn.2015.7219846

# A Brief Study on Analyzing Student's Emotions with the Help of Educational Data Mining


**S. Aruna, J. Sasanka, and D. A. Vinay**

**Abstract** Recently, the idea has reached toward considering the emotions in the learning procedure which prompts to design an innovative framework that empowers correlative analysis and classifications of various emotional variations of an individual. There is an absence of teaching pedagogues to distinguish the standard articulations of people. In the past decade, research articles have recorded the lacking properties and attempt to recognize the equivalent benefit to overcome the difficulties. This enhances the part of identifying emotions as a device to perceive the sentiments of understudies while learning. This article encompasses the record of all analytical examinations of student's emotions by applying different strategies, models, calculations and devices. This article wraps the considered works and gives the examination of qualities, shortcomings, openings, whose parts are addressed and the future work to be accomplished.

**Keywords** Emotion · Educational data mining · Data mining · Students · Educational · Students emotions

## 1 Introduction

'Emotion' is for the most part insinuated as a state of mind of a person, which has been distinctively involved with slants, consideration, and social retaliation. Emotion produces anatomical and psychological alterations. Initially, emotions or the state of feelings toward a certain occurrence of events proposed to empower reasonable practices reliant on past experiences.

At the present time, feelings impact our life decisions consistently with or without our understanding. It has been watched every once in a while that most of the part fail at disguising our sentiments. Even though, it is not possible to present our emotions explicitly to establish a connection of a specific inclination toward a feeling through

S. Aruna (✉) · J. Sasanka · D. A. Vinay
IT Department, Vasavi College of Engineering(Autonomous), Hyderabad, Telangana, India
e-mail: s.aruna@staff.vce.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_61

our activities. The outward appearance and non-verbal communication moreover will in general be an important strategy for delineating our inclination explicitly without knowing.

Learning style basically applies to the way a person tries to learn a specific topic, thereby resulting in gaining knowledge. Hence, various individuals try to grasp in unique ways. Although people may have a combination of non-similar ways of learning, still some have a prevalent style of getting the hang of contingent upon the conditions while others may have an alternative learning style. Data mining refers to a computing technique that enables the user to extract valuable information from a pool of data, a lot of which might be irrelevant to the user. Nowadays, the use of DM in the education arena is incipient and gives birth to the educational data mining (EDM) research field [1].

Educational data mining (EDM) enables paradigm-oriented learning which is used to draft models, methods and approaches to explore context information of educational amends. The increase of technology use in educational systems has led to the storage of large amounts of student data, which makes it important to use EDM to improve teaching and learning processes [2, 3]. EDM is useful in many different areas including identifying student emotions, identifying priority learning needs for different groups of students. By analyzing student emotions may reveal a relationship between a student's grade in a particular course and the interest the student has in that particular course. It discovers examples and makes expectations that describe students' practices and accomplishments, area information content, evaluations, enlightening functionalities and applications [4, 5].

In the beginning, it looks at the course of action of EDM to learn and set the informational condition as demonstrated by the learner's profile before teaching to a class. In the initial stage where the student interacts with the system, it would be beneficial if the EDM secures log information and breaks down their importance to propose recommendations. The following stage requires the EDM to evaluate the given training data which can be, for instance, the comprehensibility of the predictions.

The layout of the article is organized into three segments. Section 1 includes the introduction. Section 2 describes the supported survey of relating research along with the tools used, and finally, Sect. 3 concludes the research work.

## 2 Literature Survey

Table 1 summarizes the various survey articles observed for recognizing the student emotion, whose concepts have been incorporated into our study.

Table 2 shows the various tools applied in a student emotion recognition system.

**Table 1** Summary of various survey

| S. No. | Title | Author name | Summary of the paper |
|---|---|---|---|
| 1 | Analysis of students emotion for Twitter data using naive Bayes and nonlinear support vector machine approach [6] | Ranjeeta Rana, Mrs. Vaishali Kolhe | Analysis on samples taken from tweets related to engineering students' college life is conducted. The proposed work is to explore engineering student's informal conversations on Twitter in order to understand issues and problems students encounter in their learning experiences. To classify tweets reflecting students' problems multi-label classification algorithms is implemented. Nonlinear support vector machine, naive Bayes and linear support vector machine methods are used as multi-label classifiers which are implemented and compared in terms of accuracy. The comparison result has shown more accuracy of nonlinear SVM than naive Bayes classifier and linear SVM. Future work could analyze students generated content others than texts (e.g. images and videos), on social media sites other than Twitter (e.g. Facebook Tumbler and YouTube) |

(continued)

**Table 1** (continued)

| S. No. | Title | Author name | Summary of the paper |
|---|---|---|---|
| 2 | Toward building a computer-aided education system for special students using wearable sensor technologies [7] | Raja Majid Mehmood and Hyo Jong Lee Kusum Yadav, Fayez Alvarez | The selected EEG device headset certainly has several advantages over traditional EEG systems, being less expensive, convenient and easy to access. ModernEEG-based computer-aided systems may help develop the students' abilities by boosting learning, thinking communication skills and cooperation<br>In this paper, it gives feedback to instructors which guide further treatment in case of any mood disruption in students. The instructor-on-duty may use the previous mood history and treatment guidelines of current subjects in case of any mood disruption problem. This system adopted the IAPS protocol based on academic emotions for appropriate decision-making and treatment by instructor |
| 3 | A global perspective on an emotional learning model proposal [8] | Ana Raquel Faria, Ana Almeida, Constantino Martins, Ramiro Goncalves | The main goal of this paper was to find out if the consideration of student's emotional state influences their learning process. An emotional learning model was described, and a software prototype was developed and tested. To evaluate the prototype, two pre-tests and a final test were conducted. This paper did not address long-term personal problems might have on learning process |

(continued)

**Table 1** (continued)

| S. No. | Title | Author name | Summary of the paper |
|---|---|---|---|
| 4 | Prediction of sentiment analysis on educational data based on deep learning approach [9] | Ms. Jabeen Sultana, Ms. Nasreen Sultana, Dr. Kusum Yadav, Fayez Alvarez | This paper proposes a model based on deep learning approach to perform sentiment analysis on educational data (online courses). In this paper, we focused on the accuracy and performance of the training data set to predict the best model. MLP and SVM are recognized as the outperforming models. It is observed that SVM and MLP-deep learning methods perform well |

(continued)

**Table 1** (continued)

| S. No. | Title | Author name | Summary of the paper |
|---|---|---|---|
| 5 | Emotional strategy in the classroom based on the application of new technologies: an initial contribution [10] | Chinua Boonroungrut | In this paper, it intended to avoid boredom in class with the introduction of EVA and determine if levels of attention based on positive emotions are maintained for a considerable time in which it is included in the teaching–learning process, theory, recommendations and evaluations. In this study, they show a learning mechanism based on the inclusion of NTICS based on our EVA robot and implemented using tools SentiStrength is a short text classifier. Google's Vision API allows to detect individual objects and faces within the images, is a software that allows you to identify emotions based on images and text classification automatically with the use of the library of text mining, intercalation, and naive Bayes is used to classify the text. Limitations of this paper it is clarified that the discourse of the teacher and students is not divided, so in future investigations an individual analysis will be carried out in order to particularize the emotional |

(continued)

**Table 1** (continued)

| S. No. | Title | Author name | Summary of the paper |
|---|---|---|---|
| 6 | Exploring classroom emotion with cloud-based facial recognizer in the Chinese beginning class: a preliminary study [11] | Chinua Boonroungrut, Kim One | The facial emotion recognizer (FER) detection technology in the education field is in the early stage. Objective is to investigate the classroom emotion and the effectiveness of the Microsoft cloud-based FER interpretations The randomly selected 29 international students who enrolled the fundamental Chinese language course were investigated during five study weeks using the paper-based student outcome survey which measured teaching, assessment, generic skill and learning experience. The students' overall outcomes were over average. FER indicated that the neutral emotion was the highest detected score. The APIs reliability was still considerably questioned. Using updated quality and technology of photograph and video recording were recommended in the further research |
| 7 | Survey on analysis of students' emotions through social media data mining [12] | Ranjeeta Rana, Mrs. V.L. Kolhe | A multi-label classification algorithms to classify tweets reflecting students' problems is implemented |
| 8 | A survey and a data mining-based analysis of recent works [13] | Alejandro Peña-Ayala | The review concludes with a snapshot of the surveyed EDM works and provides an analysis of the EDM strengths, weakness, opportunities and threats, whose factors represent, in a sense, future work to be fulfilled |

(continued)

**Table 1** (continued)

| S. No. | Title | Author name | Summary of the paper |
|---|---|---|---|
| 9 | Understanding student academic achievement emotions toward business analytic course: a case study among business management students from India [14] | K. Jena | A case study among business management students from India suggests that positive emotions have a positive impact on student learning and negative emotions like boredom hinders the learning process. It also talks about the role of gender of a student in determining their degree of emotions toward the course which is business analytics, the paper proposes the use of a questionnaire |

**Table 2** Tools utilized for student emotion recognition system

| S. No | Name of the paper | Tools used |
|---|---|---|
| 1 | Correlation of student's precursor emotion toward learning science interest using EEG [15] | 1. Electroencephalogram (EEG) to study on precursor emotion effects<br>2. Nuprep electro-gel was used to clean the scalp surface<br>3. Ten20TM conductive<br>4. Features will then be extracted using the mel-frequency cepstral coefficient (MFCC) method |
| 2 | Toward emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches [16] | 1. Logitech webcam software<br>2. Microsoft Kinect for windows was used—a face-tracking engine that processes the image captured by the device and tracks human faces<br>3. Face-tracking SDK's face-tracking engine |
| 3 | Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning [17] | 1. FACEAPI—a commercial function library for tracking and locating facial landmarks, was used for frame-by-frame capture of facial coordinates, which are independent of head pose |
| 4 | Student learning context analysis by emotional intelligence with data mining toots [18] | 1. Kappa statistic<br>2. Mean absolute error<br>3. Root mean squared error<br>4. Relative absolute error |
| 5 | Analysis of students emotion for Twitter data using naive Bayes and nonlinear support vector machine approach [6] | 1. Amazon mechanical turk to create data sets for ML models<br>2. Confusion matrix for determining the accuracy of the various algorithms |
| 6 | Prediction of sentiment analysis on educational data based on deep learning approach [9] | 1. Educational dataset taken from the Kalboard 360 dataset<br>2. WEKA—This tool provides methods for a whole range of data mining tasks like data pre-processing classification, clustering, association and visualization<br>3. Confusion matrix for determining the accuracy of the various algorithms |

(continued)

**Table 2** (continued)

| S. No | Name of the paper | Tools used |
|---|---|---|
| 7 | Understanding student academic achievement emotions toward business analytic course: a case study among business management students from India [14] | 1. Statistical package for social science (SPSS-23) software tool |
| 8 | Toward building a computer-aided education system for special students using wearable sensor technologies [7] | 1. TrueScan32—an EEG device<br>2. Bluetooth connection adaptor (BAC)<br>3. Emotion recognition module (ERM) |
| 9 | Emotional strategy in the classroom based on the application of new technologies: an initial contribution [10] | 1. SentiStrength—predicts the positive and negative feelings of texts simultaneously which assigns a value scale to the feeling<br>2. Google Ajax API—allows you to add Google search results to a website through JavaScript allowing you to display the search results easily and dynamically<br>3. Google's Vision API—allows you to detect individual objects and faces within the images<br>4. HER—is a software that allows you to identify emotions based on images and text classification |

## 3 Conclusion

In this article, the insightful investigation of learner's feelings through educational information mining was discussed. The problems regarding learner's emotions are adjuring to be addressed have been investigated and examined, the models applied to depict various feelings, the classification procedures and various tools utilized in past research, examining the given work, its qualities and shortcomings are considered. The future degree drawn from the examination can be fused.

The use of various techniques of data mining such as the multi-layer perceptron (MLP) was suggested. The accuracies of the various concepts and tools of emotions in the aforementioned field and depending upon the metric were obtained, and the most optimal technique/algorithm to accomplish the desired outcomes and results of this research work were recommended.

## References

1. Baker RS, Inventado PS (2014) Educational data mining and learning analytics
2. Heubner RA, Richard A (2013) A survey of educational data mining research
3. Sachin RB, Vijay MS (2012) A survey and future vision of data mining in educational field
4. Gulhane Y, Ladhake SA (2019) Stress analysis using speech signal. Springer Nature Singapore Pte Ltd
5. Anjewierden A, Kolloffel B, Hulshof C (2007) Towards educational data mining: using data mining methods for automated chat analysis to understand and support inquiry learning processes. In: Proceedings of the international workshop on applying data mining in eLearning, pp 23–32
6. Rana R, Kolhe V (2015) Analysis of students emotion for Twitter data using Naïve Bayes and non-linear support vector machine approaches
7. Mehmood RM, Lee HJ (2017) Towards building a computer-aided education system for special students using wearable sensor technologies
8. Faria AR, Almeida A, Martins C, Gonçalves R (2017) A global perspective on an emotional learning model proposal
9. Sultana J, Sultana N, Yadav K, AlFayez F (2018) Prediction of sentiment analysis on educational data based on deep learning approach
10. Boonroungrut C (2019) Emotional strategy in the classroom based on the application of new technologies: an initial contribution
11. Boonroungrut C, One K (2019) Exploring classroom emotion with cloud-based facial recognizer in the chinese beginning class: a preliminary study
12. Rana R, Kolhe VL (2014) Survey on analysis of students' emotions through social media data mining
13. Peña-Ayala A (2014) A survey and a data mining-based analysis of recent works
14. Jena K (2019) Understanding student academic achievement emotions towards business analytic course: a case study among business management students from India
15. Nor NM, Bar AW, Salleh SHS (2015) Correlation of student's precursor emotion towards learning science interest using EEG
16. Saneiro M, Santos OC, Salmeron-Majadas S, Boticario JG (2014) Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches

17. Chu H-C, Tsai WW-J, Liao M-J, Chen Y-M (2017) Facial emotion recognition with transition detection for students with high-functioning autism in adaptive e-learning
18. James SP, Ramasubramanian P, Angeline DMD (2017) Student learning context analysis by emotional intelligence with data mining tools

# IoT-PSKTS: Public and Secret Key with Token Sharing Algorithm to Prevent Keys Leakages in IoT


**K. Pradeepa and M. Parveen**

**Abstract** In the Internet of Things (IoT) paradigm, devices involved will frequently face different types of attacks, for example flood attacks, eavesdropping attacks and so on. Once attackers have compromised the IoT device, the data materials of the IoT device will not remain confidential, and it will then be captured by the attacker; this will in turn threaten the entire network. Consequently, to safeguard IoTs, this paper proposes a public and secret key with token sharing (IoT-PSKTS) algorithm to avoid key leakages in IoT. Cryptography can be utilized for secure communication in the presence of attackers. In cryptography, a conventional public key cryptosystem will be suitable because they do not require the sender and receiver to deliver the same secret to contact without threat. But, they regularly depend on intricate mathematical calculations and are thus much more incompetent than comparable symmetric key cryptosystems. In a large number of applications, the high price of encrypting lengthy messages in public key cryptography can be prohibitive. A hybrid system tackles it by utilizing a mixture of both. In IoT network, admin generates a public key, private key, secret key and token. The public and secret key will be used for packet encryption in IoT devices and base station side, and the private key will be utilized for decryption in the admin side and token used for IoT devices access control. For encryption purpose, admin shares public and secret key with a token for IoT devices and base station. Therefore, PSKTS algorithm has been used to securely share the public and secret key with a token for IoT devices and base station in a distributed way. The experimental results show that the proposed PSKTS algorithm shares a public and secret key with token in a secured way.

**Keywords** Access control · Cryptography · Hybrid system · Packet encryption · Decryption · Secret sharing

K. Pradeepa (✉) · M. Parveen
Cauvery College for Women (Autonomous) (Affiliated to Bharathidasan University), Trichy, India
e-mail: pradeepa.cs@cauverycollege.ac.in

M. Parveen
e-mail: parveen.it@cauverycollege.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_62

## 1 Introduction

Internet of Things (IoT) is a novel technology in the fields of science, engineering, manufacturing and policy. It has fascinated the significant attention of innovation, introduction and improvement proceedings as well as is a well-known topic of both social media and press. The significance of IoT is the hopeful alteration of features, for daily life proceedings and services. In the IoT, devices are mostly subject to various kinds of attacks, i.e. flood attacks, eavesdropping attacks and so on. Once attackers have compromised an IoT device, IoT device information items are nothing confidential and captured by attackers; thus, the whole network menaced. To address this issue, cryptography is required.

Cryptography is a way of securing data and communication via coding so that only those who targeted can read and process it [1]. Cryptographic algorithms utilize a set of encryption and decryption processes to protect communication between IoT devices, the IoT base station and the admin. The cypher suite uses one encryption algorithm, another authentication algorithm and another key swap algorithm. This process embedded in protocols and encoded in software that works on networked computer systems, and operating systems include public and private key generation for information encryption/decryption, digital signature and message verification and key exchange.

In cryptography, conventional public key cryptography is suitable because they do not require the sender and receiver to share the same secret to contact safely. But, they usually depend on complex mathematical calculations and are therefore often doing not work much better than symmetric key cryptography. For most applications, the high price of encrypting lengthy messages in public key cryptography may be prohibitive. Hybrid cryptography deals it by utilizing a mixture of both. Many previous security programs assume that no data about the secret key used to encrypt information during the encryption procedure will leak. But, practically, this scenario is not surely true. Data regarding the secret key could compromise for different attacks, for example side-channel attacks on the communication cable, physical memory, etc.

Consequently, a cryptographic algorithm that has theoretically proven to be very secure will be practically completely unsafe if implemented improperly. Let us take the side-channel attack as an example. By measuring the power consumption and computation time, the attacker can get the internal state of an intermediate physical device. The attacker uses this data to break a cryptographic primitive computed by the device. The attacker uses “cold boot” attack, for steal an important fraction of the bits of a cryptographic key if the key recorded in memory while key sharing. This problem motivates to generate well secured key sharing algorithm. The main objective of this paper is to prevent key leakages in IoT to enhance security.

Also, access control is a way to ensure that IoT devices are the ones they claim to be and that they have proper packet access [2]. Take two IoT devices, for example. The two IoT devices are side by side. Also, the two IoT devices joined into different

IoT networks. Each IoT network is monitored individually by each IoT base station. If there is no access control at this time, the information going to one IoT device is likely to be accessed by another IoT device. Access control becomes essential to avoid this. Token-based access control is an authentication technique that provides extra protection. Each IoT device contains a token using this technique. Without this token, it is unfeasible to use the packet. Also, secret sharing means sharing a secret between collections of IoT devices, each of whom allotted a share of the secret [3]. The secret could redo by only the proper IoT device and when enough number, or perhaps various types, of shares, merged. In this work to safely share keys to all IoT devices for encryption, this paper proposes a public and secret key with token sharing (IoT-PSKTS) algorithm to avert key leakages. This algorithm assists in admin for the distribution of keys safely.

The rest of the paper is organized as follows: Sect. 2 presents a related work of existing cryptography, access control and secret sharing techniques. Section 3 describes a public and secret key with token sharing algorithm to prevent keys leakages in detailed. Section 4 provides extensive experimental results to support the proposed PSKTS algorithm. Finally, Sect. 5 provides the conclusion of the work.

## 2 Related Work

This section discusses related works of previous cryptography techniques, access control algorithms and secret sharing techniques.

Carracedo et al. [1] provided an in-depth cryptography research study related to IoT safety. The authors discussed security improvements from cryptography research in response to security challenges posed by the IoT environment. By separating cryptography research, a collection of conclusions and suggestions for future research guidelines was provided.

Alramadhan and Sha [2] initially provided a comprehensive survey of previous access control systems and examined their effectiveness on IoT networks. Subsequently, both challenges in building IoT access control and future IoT access control management objectives are recognized and debated.

Tang et al. [3] presented a cost-effective secret key distribution system for multiple-input multiple-output (MIMO) IoT applications, which can detect the secret key distribution and simultaneous communication. The secret key data joined to a compilation of small active/inactive channels of the legitimate recipient, generated by single value corruption (SVC) MIMO channel of the legitimate recipient. The legitimate recipient can precisely locate the shared key and delete down the contact data with the discretionary low key BER and bit error rate (BER).

Yousefi and Jameii [4] have developed a hybrid encryption algorithm that was developed to decrease security threats and improve encryption speed and a smaller amount of computational complicated. The motive of this hybrid algorithm is the reliability of the information, privacy, non-rejection in data transmission for IoT.

Finally, the proposed encryption algorithm was developed by the MATLAB software, and its speed and security performance wastested compared to the standard encryption algorithm.

Gunathilake et al. [5] presented the need of lightweight cryptography (LWC), its present position, related protocols and technologies, i.e. LoRaWAN, and the challenges in the current circumstance by examining previous practical and theoretical studies in the curriculum.

Terkawi et al. [6] discussed models, issues and challenges raised by IoT. Also, the authors discussed different types of access controls. The standard access control models will eventually transform into something very different and more relevant to the Internet of Things, such as their proposed model combining authentication and verification. Furthermore, implementing rules will enable greater flexibility and enable decentralized management of shared resources, which is a significant goal in a certified joint environment.

Wang et al. [7] designed a reliable attribute-based access control model, which introduced the confidence rating module and passed recognition with artificial control, combining dynamic confidence attribute and multiple standard attributes. Finally, their model has very advanced features and improves the security of the IoT system with test results.

Surendran et al. [8] initially debated the requirement for IoT cryptography and their plan variations with standard block cyphers. A survey of other IoT cryptographic algorithms debated later. Also, they looked at the various kinds of attacks that researched in some of these cyphers. Lastly, they compare the effectiveness of some of these cyphers on embedded and Windows platforms.

Miao and Jiang [9] introduced a method of sharing secrets between two users using the effects of wireless channel dynamics on the data connection layer. In particular, the authors create problems in improving their goal of minimizing the eavesdropper's access to all secret sharing packets. Their contributions are: (i) authors coming up with a secret sharing method that reduces the objective function mentioned above and (ii) authors analyzing how authors find the worst chance of eavesdropper accessing all the secretly distributed packets. Simulations verify their theoretical outcomes.

Farhadi et al. [10] discussed a new way of storing integrated data in IoTs via secret sharing scheme using $(t, n)$-threshold in cloud storage. In this way, the actual data is divided into t blocks, where each block is treated as a share. This method is scalable and detectable; that is, it can insert new data or delete a portion of the actual data, without changing shares, and when cloud service providers find fault with sending the wrong share. Table 1 shows the related work comparison based on the security type used.

**Table 1** Comparison of related work based on the type of security used

| Related work | Type of security used |
|---|---|
| [1] | Cryptography |
| [2] | Access control |
| [3] | Secret sharing |
| [4] | Cryptography |
| [5] | Cryptography |
| [6] | Access control |
| [7] | Access control |
| [8] | Cryptography |
| [9] | Secret sharing |
| [10] | Secret sharing |

## 3 IoT-PSKTS: Public and Secret Key with Token Sharing Algorithm to Prevent Keys Leakages

IoT network consists of three entities: (1) Admin, (2) IoT base station and (3) IoT devices. Admin is a person who responsible for forest monitoring; he can monitor forest anywhere at any time using IoT. IoT devices monitor forest and generate sensed data. Followed by, it shares sensed data to the admin through IoT base station. IoT base station is a network controller of IoT. After collection of sensed data, it outsources these sensed data to admin. IoT devices have limited energy, limited coverage area, limited transmission power (Et) and limited receiving power (Er). Each IoT devices is fixed at various locations in the forest for sensing temperature, sound, vibration, pressure, etc. This section proposed IoT-PSKTS algorithm to improve IoT network security, which is explained in Algorithm 1, and the IoT-PSKTS algorithm block diagram is shown in Fig. 1.

Demonstrated in Fig. 1, the admin generates three keys with one token to all IoT devices and base station for encryption, decryption and access control purpose (Step 2–4). They are,

1. The public key for IoT base station (for encryption) with
2. The secret key for each IoT devices (for encryption) and
3. Token for each IoT devices (for access control)
4. The private key (for decryption) for own use.

Followed by that, the admin converts public key, secret keys and tokens to $n$ number of shares (Steps 5–7) and then forwards these $n$ shares to IoT base station (Step 8). After receiving any number of shares ($<n$), the IoT base station reconstruct public key, secret keys and tokens based on shares (Steps 11–13). Followed by that, the IoT base station converts secret keys and tokens of each IoT devices to $n$ number of shares (Steps 16–18). Each share will transmit to the IoT device, which is entitled

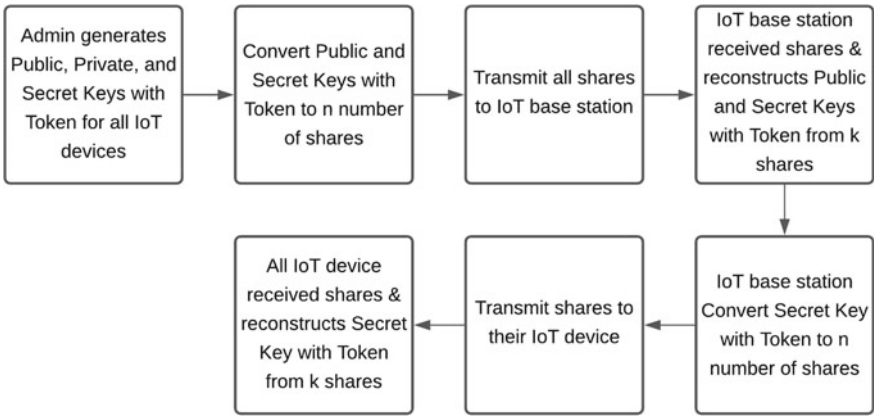



**Fig. 1** Block diagram of IoT-PSKTS algorithm

to it (Step 19). After receiving any number of shares ($<n$), all IoT devices reconstruct its own secret key and token, respectively (Step 21–25).

**Algorithm 1: IoT-PSKTS: Public and secret key with token sharing algorithm**

| | | |
|---|---|---|
| **Input** | : | Admin, IoT base station (IBS), IoT devices (IDSs) |
| **Output** | : | The public and secret key with token sharing |
| | | **Admin Side** |
| **Step 1** | : | For each IoT device ID from IDSs do |
| **Step 2** | : | Keys[] = Key_Generation() // **Algorithm 2** |
| **Step 3** | : | PubKey = Keys[0], PrivKey = Keys[1], SecKey = Keys[2] |
| **Step 4** | : | Token = Generate_Token(ID) // **Algorithm 3** |
| **Step 5** | : | AK = PubKey + "||" + SecKey + "||" + Token; |
| **Step 6** | : | Let N, k // N - no of shares generated & k - no of shares must for reconstruct |
| **Step 7** | : | SH[] = Generate_Shares(AK, N, k) //**Algorithm 4** |
| **Step 8** | : | Forward SH with k to IBS |
| **Step 9** | : | End For |
| | | **IoT Base Station Side** |
| **Step 10** | : | For each IoT device ID from IDSs do |
| **Step 11** | : | IBS received SH with a k value |
| **Step 12** | : | Take k shares for reconstructing |
| **Step 13** | : | AK = Reconstruct(k shares) // **Algorithm 5** |
| **Step 14** | : | Keys[] = AK.split("||"); |
| **Step 15** | : | PubKey = Keys[0], SecKey = Keys[1], Token = Keys[2] |
| **Step 16** | : | AK = SecKey + "||" + Token; |
| **Step 17** | : | Let N, k // N - no of shares to be generated & k - no of shares must for reconstruct |
| **Step 18** | : | SH[] = Generate_Shares(AK, N, k)// **Algorithm 4** |
| **Step 19** | : | Forward SH with k to ID |
| **Step 20** | : | End For |
| | | **IoT Device Side** |
| **Step 21** | : | ID received SH with a k value |
| **Step 22** | : | Take k shares for reconstructing |
| **Step 23** | : | AK = Reconstruct(k shares) // **Algorithm 5** |
| **Step 24** | : | Keys[] = AK.split("||"); |
| **Step 25** | : | SecKey = Keys[0], Token = Keys[1] |

### 3.1 Key Generation

Cryptography techniques are used to convert the original text message into unreadable text format (ciphertext). Three different types of keys such as private key, public key and secret key are used in the hybrid cryptography. In encryption, public and secret keys are involved, and in decryption, private and secret keys are involved. Algorithm 2 explained the process of key generation algorithm by using two different prime numbers, and it is denoted as $p1$ and $p2$. Step 2 multiples both prime numbers. Step 3 will subtract from prime number and multiplies it with $m2$. Step 4 calculates the co-prime number which contains common factor between two prime numbers.

$$35 = 7 \times 5 \times 1$$
$$39 = 13 \times 3 \times 1 \quad (1)$$

In mod 7, the multiplicative inverse of 2 is 4. This algorithm calculates the multiplicative inverse of $e1$ with mod $m2$ (Step 5) which provides $e2$. The values $e1$ and $m1$ are taken for the public key (Step 9), and also, $e2$ and $m1$ are taken for the private key (Step 10). The AES algorithm is used to generate a secret key. Firstly, this algorithm produces an AES key generator (Step 6) which provides the functionality of a secret (symmetric) key generator. Followed by this, the algorithm launches this key generator with a 128-bit key size (Step 7). Also, this key generator creates a secret (Step 8). Following this, this algorithm encodes this secret into the secret key (Step 11).

### 3.2 Token Generation

This section creates a token using the HMAC-SHA-1 algorithm because HMAC-SHA-1 algorithm provides a strong message authentication code based on SHA1, which is a cryptographic hash function over the data (to be authenticated) and a secret key [11, 12]. This message authentication code is used as a token. This token is used to control unauthorized access. In this case, IoT device identity is used as data and HMACSHA1 keyword is used as the secret key.

Algorithm 3 describes the making of tokens.

To make HMAC-SHA-1 token over the IoT Device Id data, the following are executed,

1. SecretKeySpec class—This class defines the secret key in the provider-independent fashion (Step 1). It can use to build SecretKey from a byte array, without having to go through (based on the provider) SecretKeyFactory. This class is for use with raw secret keys that can be displayed as a byte array and have no key parameters associated with them, e.g. DES or triple-DES keys.

| **Algorithm 2: Key_Generation** | | |
|---|---|---|
| **Input** | **:** | Two Random Prime numbers (p1 and p2) |
| **Output** | **:** | Keys[] |
| **Step 1** | **:** | Let Keys[] = {} |
| **Step 2** | **:** | m1 = p1 * p2 |
| **Step 3** | **:** | m2 = (p1-1) * (p2-1) |
| **Step 4** | **:** | e1 = getCoprime(m2) |
| **Step 5** | **:** | e2 = modInverse(e1, m2) |
| **Step 6** | **:** | keyGen = KeyGenerator.getInstance("AES") // Secret key generation based on AES algorithm |
| **Step 7** | **:** | keyGen.init(128) |
| **Step 8** | **:** | secret = keyGen.generateKey() |
| **Step 9** | **:** | PubKey = e1, m1 |
| **Step 10** | **:** | PrivKey = e2, m1 |
| **Step 11** | **:** | SecKey = Base64.encode(secret.getEncoded()) |
| **Step 12** | **:** | Keys[0] = PubKey, Keys[1] = PrivKey, Keys[2] = SecKey |
| **Step 13** | **:** | return Keys[] |

2. Mac class—This class provides the operation of the "Message Authentication Code" (MAC) algorithm (Step 2). The Mac.getInstance (HMAC_SHA1_ALGORITHM) method returns the Mac object that executes the default MAC algorithm.
3. Mac object with a given key started in Step 3.
4. mac.doFinal (data.getBytes ()) method processes a given array of bytes (rawHMAC) and terminates MAC operation (Step 4).
5. Lastly, this algorithm encodes raw HMAC to token using the encode function (Step 6–Step 17).

**Algorithm 3: Generate_Token(ID)**

| | | |
|---|---|---|
| **Input** | **:** | IoT Device (ID), HmacSHA1 |
| **Output** | **:** | Token |
| **Step 1** | **:** | signingKey = new SecretKeySpec(ID.getBytes(), HmacSHA1) |
| **Step 2** | **:** | mac = Mac.getInstance(HmacSHA1) |
| **Step 3** | **:** | mac.init(signingKey) |
| **Step 4** | **:** | rawHmac[] = mac.doFinal(ID.getBytes()) |
| **Step 5** | **:** | Token = new String(encode(rawHmac)) |
| **Step 6** | **:** | return Token |
| | | **encode function** |
| **Step 7** | **:** | char[] encode(byte[] bytes) |
| **Step 8** | **:** | { |
| **Step 9** | **:** | char[] HEX = {'0', '1', '2', '3', '4', '5', '6', '7', '8', '9', 'a', 'b', 'c', 'd', 'e', 'f'} |
| **Step 10** | **:** | amount = bytes.length; |
| **Step 11** | **:** | char[] result = new char[2 * amount] |
| **Step 12** | **:** | j = 0 |
| **Step 13** | **:** | For (int i = 0; i < amount; i++) |
| **Step 14** | **:** | result[j++] = HEX[(0xF0 & bytes[i]) >>> 4] |
| **Step 15** | **:** | result[j++] = HEX[(0x0F & bytes[i])] |
| **Step 16** | **:** | End For |
| **Step 17** | **:** | return result |
| **Step 18** | **:** | } |

### 3.3 Shares Generation

After key generation, the admin wants to send the public key, the secret key with a token to IoT base station. Furthermore, the IoT base station wants to send the secret key with a token to each IoT devices. But, key sharing in an IoT network is computationally unsafe and inconsistent due to its dynamic nature. The secure key sharing technique is necessary to tackle this problem. Before generating shares, all keys should be merged (AK). Then, this technique splits the merged keys into many shares and then transmits each share to the IoT base station or IoT devices. Shares generation is explained in Algorithm 4 and Fig. 2.

This algorithm takes AK, $N$ and $k$ for input (Step 1). Here,

$N$—Number of shares to be generated and

$k$—Number of shares must reconstruct.

$$f(x) = a0 + (a1 * x) + \left(a2 * x^2\right) \tag{2}$$

where $a0$ is the secret, and $a1$ and $a2$ are randomly chosen integers. AK converted to $N$ shares is shown in Fig. 2 clearly.

| **Algorithm 4: Generate_Shares(AK, N, k)** | | |
|---|---|---|
| **Input** | : | AK, N, k |
| **Output** | : | Shares (SH) |
| **Step 1** | : | Let a0 = AK, SH = {} |
| **Step 2** | : | Generate Random (k-1) numbers (a1 and a2) |
| **Step 3** | : | Let f(x) = a0 + (a1 * x) + (a2 * $x^2$) |
| **Step 4** | : | For x=1; x<=N; x++ |
| **Step 5** | : | SH[x-1] = (x,f(x)) |
| **Step 6** | : | End For |
| **Step 7** | : | return SH |

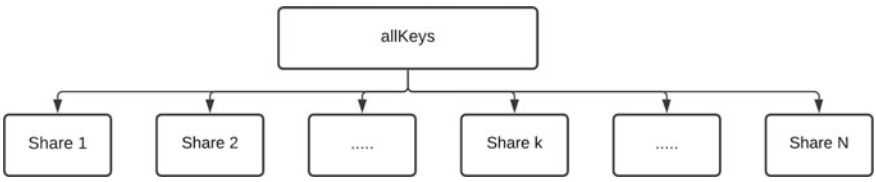



**Fig. 2** Shares generation

### 3.4 Reconstruct

After receiving all $N$ shares, an IoT base station or IoT device takes $k$ shares for reconstructing the AK. It is the advantage of the proposed IoT-PSKTS algorithm. If any intermediate device turns into malicious, it may be dropping any shares. So, this algorithm takes $k$ shares which are enough for reconstructing, which shows in Fig. 3.

Reconstruct process explained in Algorithm 5. In Fig. 3, the algorithm applies $k$ shares into a Lagrange polynomial formula (Step 1).

$$f(x) = \sum_{j=0}^{k} (yj * lj(x)) \tag{3}$$

It provides $a0$, $a1$, $a2$ and so on. For simplicity, $k$ values are 3 (it give while shares generation), so this algorithm provides $a0$, $a1$ and $a2$ only (Step 2). Here, $a0$ is AK (Step 3).

**Algorithm 5: Reconstruct(k shares)**

| | | |
|---|---|---|
| **Input** | : | k shares (x0,y0), (x1,y1), ..... , (xk,yk) |
| **Output** | : | AK |
| **Step 1** | : | f(x) = $\sum_{j=0}^{k}$(yj * lj(x)) // Lagrange basis polynomial<br>// If k=3 then f(x) = (y0*l0(x)) + (y1*l1(x)) + (y2*l2(x)) |
| **Step 2** | : | f(x) = a0 + (a1 * x) + (a2 * $x^2$) |
| **Step 3** | : | AK = a0 |

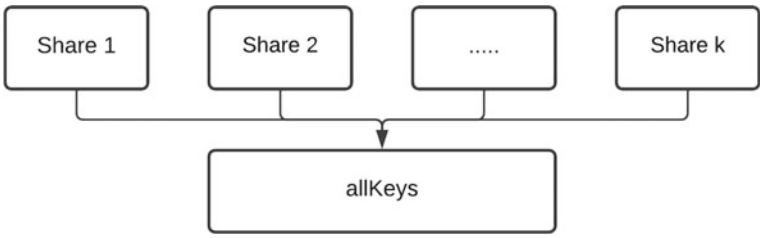



**Fig. 3** Shares reconstruction

### 3.5 Mathematical Model of Share Construction and Reconstruction Process

***Share Construction:***

For instance, AK = 1234. The proposed research work attempts to split the secret into six shares ($n = 6$), where any three shares ($k = 3$) are enough to reconstruct AK. At random, $k - 1$ numbers are obtained as: 166 and 94.

($a0 = 1234$; $a1 = 166$; $a2 = 94$), where $a0$ is secret

The polynomial mentioned in Eq. (2) provides secret shares:

$f(x) = 1234 + 166x + 94x^2$

It constructs six shares from the polynomial:

$f(1) = 1234 + 166(1) + 94(1)^2 = 1494$
$f(2) = 1234 + 166(2) + 94(2)^2 = 1942$
$f(3) = 1234 + 166(3) + 94(3)^2 = 2578$
$f(4) = 1234 + 166(4) + 94(4)^2 = 3402$
$f(5) = 1234 + 166(5) + 94(5)^2 = 4414$
$f(6) = 1234 + 166(6) + 94(6)^2 = 5614$

***Share Reconstruction:***

To reconstruct the AK, any three shares ($k = 3$) will be sufficient.

($x0, y0$) = (2, 1942); ($x1, y1$) = (4, 3402); ($x2, y2$) = (5, 4414)

Now, Lagrange basis polynomials are applied based on Eq. (3):

$l0(x) = (x - x1/x0 - x1) * (x - x2/x0 - x2) = (x - 4/2 - 4) * (x - 5/2 - 5) = x^2/6 - 3x/2 + 10/3$
$l1(x) = (x - x0/x1 - x0) * (x - x2/x1 - x2) = (x - 2/4 - 2) * (x - 5/4 - 5) = -x^2/2 + 7x/2 - 5$
$l2(x) = (x - x0/x2 - x0) * (x - x1/x2 - x1) = (x - 2/5 - 2) * (x - 4/5 - 4) = x^2/3 - 2x + 8/3$

Thus,

$$\begin{aligned} f(x) &= (y0 * l0(x)) + (y1 * l1(x)) + (y2 * l2(x)) \\ &= \left(1942 * \left(x^2/6 - 3x/2 + 10/3\right)\right) + \left(3402 * \left(-x^2/2 + 7x/2 - 5\right)\right) \\ &\quad + \left(4414 * \left(x^2/3 - 2x + 8/3\right)\right) \\ &= 1234 + 166x + 94x^2 \end{aligned}$$

Here 1234 is AK.

## 4 Results and Discussions

This section provides the results of the experimental and analysis of the IoT-PSKTS algorithm on the IoT network. This simulation assumes that most IoT devices are still distributed evenly and randomly in the forest. These IoT devices measure temperature, humidity, light intensity, wind speed, rainfall and smoke [13, 14]. Following these devices, readings are transmitted to the admin via the IoT base station. Java is used to evaluate the IoT-PSKTS algorithm. Compare the proposed IoT-PSKTS algorithm with other secret sharing algorithms, such as AdiShamir's perfect secret sharing scheme (PSS), HugoKrawczyk's computational secret sharing scheme (CSS) and Rabin's information dispersal algorithm (IDA) [15]. Table 2 shows the time take in (ms) of share formation ($n = 5, k = 3$).

Figure 4 shows the graph of time taken against data sizes in share creation when $n = 5, k = 3$.

Table 2 and Fig. 4 demonstrated that IoT-PSKTS is the fastest algorithm based on the data sizes. Furthermore, Table 3 shows time take in (ms) for share recreation ($n = 5, k = 3$).

Figure 5 shows the graph of time taken against data sizes in share recreation when $n = 5, k = 3$.

**Table 2** Time take in (ms) for share creation ($n = 5, k = 3$)

| Algorithm | Data size (in KB) | | |
|---|---|---|---|
| | 16 | 32 | 64 |
| CSS | 19.45 | 25.03 | 31.80 |
| IDA | 12.82 | 19.59 | 21.19 |
| PSS | 28.78 | 40.01 | 42.89 |
| IoT-PSKTS | 10.97 | 17.19 | 19.68 |

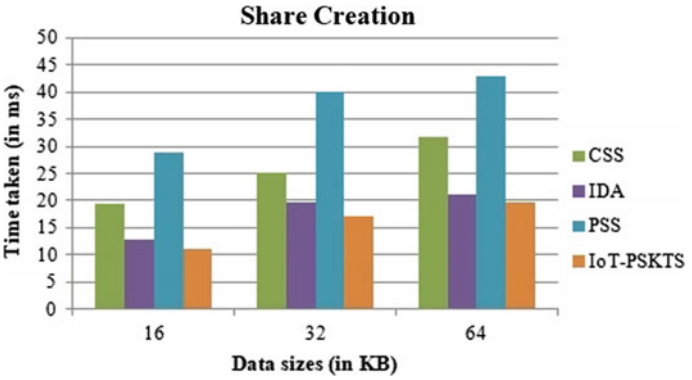



**Fig. 4** Graph of data size against the time taken in share creation when $n = 5, k = 3$

**Table 3** Time take in (ms) for share recreation ($n = 5$, $k = 3$)

| Algorithm | Data size (in KB) | | |
|---|---|---|---|
| | 16 | 32 | 64 |
| CSS | 19.27 | 21.63 | 26.11 |
| IDA | 17.82 | 19.51 | 22.57 |
| PSS | 30.01 | 20.00 | 23.89 |
| IoT-PSKTS | 15.76 | 17.04 | 20.49 |

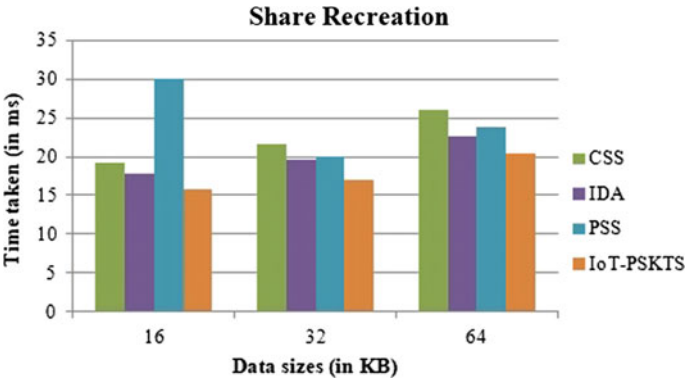



**Fig. 5** Graph of data size against the time taken in share recreation when $n = 5$, $k = 3$

Table 3 and Fig. 5 demonstrated that IoT-PSKTS is the fastest algorithm based on the data size. This approach demonstrates the data size improvement to avoid the scalability compared to other algorithms.

## 5 Conclusion

This paper proposed a public and secret key with token sharing (IoT-PSKTS) algorithm to prevent key leaks in the IoT. In an IoT network, the admin creates a public key, a private key, a secret key and a token. Here are the public and secret key used for packet encryption, the private key used for decryption and the token used to control access. For encryption purposes, the admin shares a public and secret key with the token for IoT base station and IoT devices. Consequently, the IoT-PSKTS algorithm presents security while keys shared in a distributed way. Compared with previous key sharing algorithms, experimental results have proven that IoT-PSKTS algorithms present secure key sharing with minimal share creation and recreation time in forest monitors on the IoT network.

## References


1. Carracedo JM, Milliken M, Chouhan PK, Scotney B, Lin Z (2018) Cryptography for security in Io. In: 2018 IEEE fifth international conference on internet of things: systems, management and security. ISBN: 978-1-5386-9585-2
2. Alramadhan M, Sha K (2017) An overview of access control mechanisms for the internet of thing. In: 2017 IEEE 26th International conference on computer communication and networks. ISBN: 978-1-5090-2991-4
3. Tang J, Song H, Xu A, Jiang Y, Wen H, Zhang Y, Qin K (2020) Secret sharing simultaneously on the internet of thing. In: 2020 IEEE international conference on power, intelligent computing and systems. ISBN: 978-1-7281-9874-3
4. Yousefi A, Jameii SM (2017) Improving the security of internet of things using encryption algorithm. In: 2017 IEEE international conference on IoT and application. ISBN: 978-1-5386-1698-7
5. Gunathilake NA, Buchanan WJ, Asif R (2019) Next generation lightweight cryptography for smart IoT devices: implementation, challenges and application. In: 2019 IEEE 5th world forum on internet of things. ISBN: 978-1-5386-4980-0
6. Terkawi A, Innab N, Amri SA, Amri AA (2018) Internet of things (IoT) increasing the necessity to adopt specific type of access control technique. In: 2018 IEEE 21st Saudi computer society national computer conference. ISBN: 978-1-5386-4110-1
7. Wang J, Wang H, Zhang H, Cao N (2017) Trust and attribute-based dynamic access control model for the internet of thing. In: 2017 IEEE international conference on cyber-enabled distributed computing and knowledge discovery. ISBN: 978-1-5386-2209-4
8. Surendran S, Nassef A, Beheshti BD (2018) A survey of cryptographic algorithms for IoT device. In: 2018 IEEE long Island systems, applications and technology conference. ISBN: 978-1-5386-5029-5
9. Miao L, Jiang D (2017) Optimal secret sharing for secure wireless communications in the era of internet of thing. In: 2017 IEEE 4th international conference on smart and sustainable city. ISBN: 978-1-78561-503-0
10. Farhadi M, Bypour H, Mortazavi R (2019) An efficient secret sharing-based storage system for cloud-based IoT. In: 2019 IEEE 16th international ISC conference on information security and cryptology. ISBN: 978-1-7281-4374-3
11. Fischlin M, Janson C, Mazaheri S (2018) Backdoored hash functions: immunizing HMAC and HKDF. In: 2018 IEEE 31st computer security foundations symposium (CSF). IEEE, pp 105–118
12. Suhaili SB, Watanabe T (2017) High-speed implementation of the keyed-hash message authentication code (HMAC) based on SHA-1 algorithm. Adv Sci Lett 23(11):11096–11100
13. Kishorebabu V, Sravanthi R (2020) Real-time monitoring of environmental parameters using IOT. Wireless Pers Commun 8:1–24
14. Kusuma HA, Anjasmara R, Suhendra T, Yunianto H, Nugraha S (2020) An IoT based coastal weather and air quality monitoring using GSM technology. J Phys Conf Ser 1501(1):012004
15. Buchanan WJ, Lanc D, Ukwandu E, Fan L, Russell G, Lo O (2015) The future internet: a world of secret share. Future Internet 7:445–464. https://doi.org/10.3390/fi7040445

# Investigation and Analysis of Path Evaluation for Sustainable Communication Using VANET


**D. Rajalakshmi, K. Meena, N. Vijayaraj, and G. Uganya**

**Abstract** Today, the taskforce is getting increasing and the streets are getting more hazardous by the impact of blockage and increment of crashes. Intelligent transportation systems (ITS) are utilized to incorporate data innovation in transportation. Vehicular ad hoc networks (VANETs) are a subset of MANET which provides communication between the mobile nodes. VANET is a collection of various dynamic nodes that can change and configure. In VANET, enormous routing protocols are implemented to route the packet reliably. One of the protocols is ad hoc on-demand distance vector (AODV) routing protocol, and this protocol can only be used when the nodes are in static movement. The other protocol is destination sequenced distance vector (DSDV) in which each node maintains a table of information about the presence of the other nodes. In traditional system, only one algorithm used for the communication strategy, but in our proposed system GAD protocol (a combination of Genetic, AODV and DSDV) in which the functionalities of entire algorithms can be used based on the communication range in the network. Our proposed idea, using GAD protocol (a combination of Genetic, AODV, and DSDV) in which the functionalities of both the algorithms can be used based on the communication link in the network. Here proposed a numerical model to calculate the path duration between source and destination using GAD protocol and solve the Sybil attack, its severe attack on vehicular ad hoc networks (VANETs). The numerical model is simulated, and the GAD protocol is developed using MATLAB. The result exposes that when the transmission range increased then the path duration and numbers of hops become decreased in VANET.

**Keywords** FCM · GAD · KNN · MANET · VANET

D. Rajalakshmi (✉)
Computer Science and Engineering, Sri Sairam Institute of Technology, Chennai, India

D. Rajalakshmi · K. Meena · N. Vijayaraj
Computer Science and Engineering, VelTech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

G. Uganya
Saveetha School of Engineering, Electronics and Communication Engineering, Chennai, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_63

## 1 Introduction

For humankind, transportation is the basic mode to travel from one place to another. Nowadays, the different kinds of vehicles including transports, trucks, trains, cable cars, ships, planes, bikes, bikes, and helicopters are used for transportation. In recent days, engineers play a vital role in developing software for effective communication between source and destination [1]. When a packet is traveled from source to destination, there is a chance of congestion, which means when the message traffic is so heavy it slows down the network response time [2]. This leads to an increase in delay and a decrease in performance. At the same time, it leads to a collision.

ITS is an intelligent transportation system [3]. It pointed toward utilizing PCs and interchanges to make travel more intelligent, quicker, more secure, and more advantageous. ITS makes travel more secure and less tedious and makes it simpler to choose how you wish to travel [4]. And business has a more efficient, less costly way of moving its products to the marketplace.

1. **Routing Protocol**: To choose the correct routing protocol is very important in VANETs because the topology changes frequently.
2. **Mathematical Models**: It is a mathematical expression for identifying path duration between two vehicles through an average number of hops and connection duration.
3. **Modeling Assumptions**: Our assumption is to make connection duration for the dynamic nodes.
4. **Area of Boundary Region for Identifying the Next Hop**: The node becomes closer to the boundary region, and it reduces the number of hop distance between source and destination.
5. **Probability of Identifying Nodes in the Shaded Region**: To identify the probability of a single node in the boundary region has to increase the performance of the network [5].

    1. Number of Hop: Intermediate node between source and destination.
    2. Speed of Nodes: The mandatory field of path duration is the motion and speed of the node.
    3. Connection Duration: It calculates the time for two nodes which is active and which one is directly connected to the transmission range.
    4. Path Duration: It improves the performance and throughput of VANET.

## 2 VANET

VANET is an emerging technology; it leads to improving security in wireless communication [6]. It is inherited from mobile ad hoc network (MANET). The similarities of VANETs with mobile ad hoc networks (MANETs) are flat network (infrastructureless environment), self-organizing without a server configuration [7]. VANETs can be differentiated by the following parameters: (i) extremely vibrant topology; (ii)

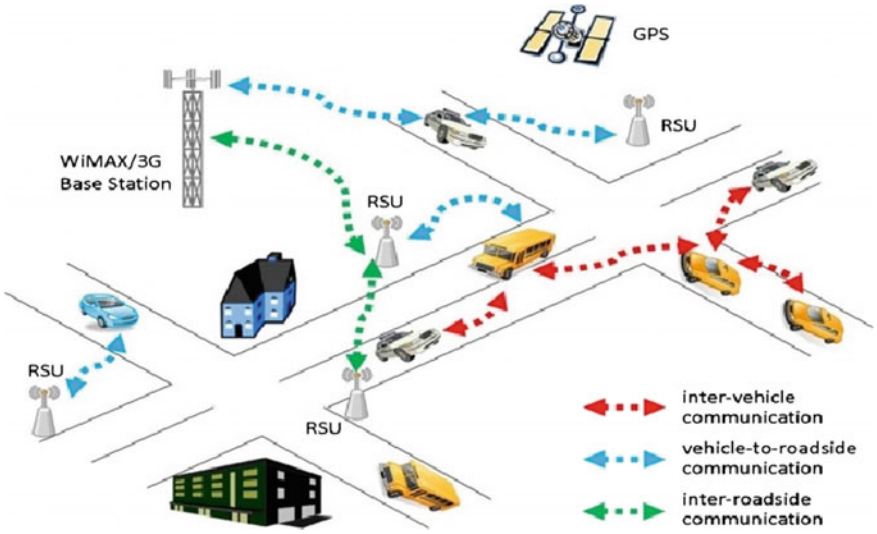



**Fig. 1** Architecture of VANET

frequently disconnected network; (iii) sufficient energy and capacity gave by the vehicle; (iv) geological kind of correspondence; (v) versatility demonstrating and forecast; (vi) hard suspension limitations; and (vii) connection with onboard sensors [8]. The architecture of VANET is shown in Fig. 1.

## 3 Related Work

In the existing system, they used AODV routing algorithms for static movement of nodes and communication with each other. DSDV routing algorithm for dynamic movement of node communicates with each other [9]. The static and dynamic movement of node communication works separately in the existing system. Before establishing communication, Initially calculated the path duration of the routes in the network, automatically performance and throughput of the VANETs can be enhanced significantly [10].

**Drawbacks in the Existing System**

1. If the path time increased, the transmission range also increased; if the transmission range is increased, then it needs more power to complete the entire communication.
2. If the number of hops increased, the average path time decreased for a fixed number of nodes.
3. If there is any data loss occurs (attack) during transmission, the packet cannot be recovered automatically.

## 4 Proposed System

In our proposed system, the communication between the nodes is established during both static and dynamic movements of nodes using GAD protocol. GAD protocol is a combination of genetic algorithm, AODV algorithm, and DSDV algorithm. The genetic algorithm identifies which protocol to be used based on node motion whether it is static, it uses the AODV routing algorithm, or its dynamic, it uses the DSDV routing algorithm helps in deciding which protocol to be used based on node motion whether it is static it uses the AODV routing algorithm, or dynamically it uses the DSDV routing algorithm. Here, the GAD protocol is used to detect and resolve the Sybil attack and successfully transmit the packet to the other nodes.

Sybil Attack: It is a basic attack in VANET. In this attack, the aggressor sends various messages to different vehicles. Each message contains an alternate source identity. It confuses for different vehicles by sending incorrect messages like congestion messages. If there is congestion, further vehicles are compelled to take another path. The primary point of the attacker is to give an illusion of various vehicles to different vehicles so vehicles can pick another course. With the help of GAD protocol, it detects the sybil attack, without disturbing the communication it resolves the attack more effectively going to detect and resolve these attacks.

**Advantages of the Proposed System**

1. Simulation results depict that the identification of reliable path time is very low for all circumstances.
2. The communication range is increased, and then the path time is low.
3. Genetic algorithm provides reliable communication and reduces path failure.

In our proposed work, the implementation was divided into three sections:

1. Section 1 explains the concept of creation of nodes and finds the best route to communicate in both static and dynamic movements of vehicles separately.
2. Section 2 includes the combination of both static and dynamic movements of vehicles using the GAD protocol and with simulation result and graph.
3. Section 3 depicts the concept of how a Sybil attack occurred and recovered.

## 5 GAD Protocol

GAD stands for genetic AODV DSDV algorithm. Generally, genetic algorithms are used to find the reliable path between the nodes. In the GAD protocol, the genetic algorithm chooses the functionality of the AODV algorithm if the nodes are connected in a static link; otherwise, it chooses the functionality of the DSDV algorithm if the nodes are connected in a dynamic link. For clustering to the nearest nodes, the fuzzy C-means (FCM) algorithm is used when AODV is selected and K-nearest neighbor (KNN) algorithm is used when DSDV is selected. Using GAD protocol, the functionalities of both the algorithms can be used based on the link in the network.

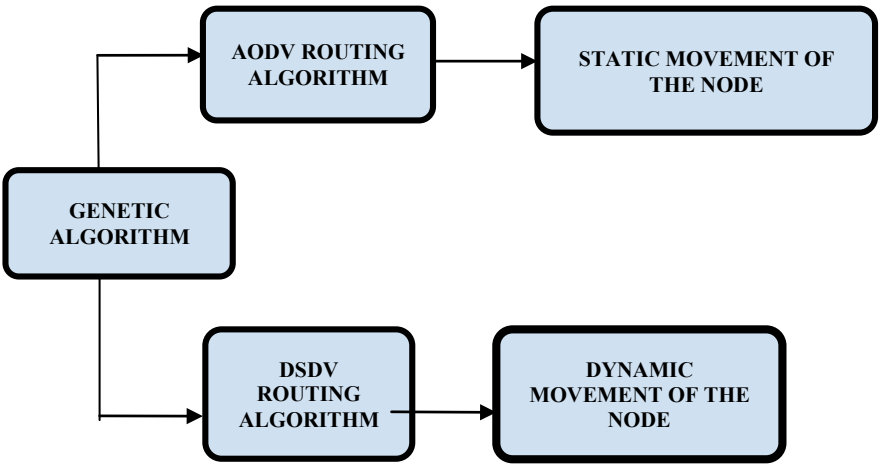



**Fig. 2** System architecture

In the GAD protocol, the genetic algorithm decides which algorithm to be used. The proposed system architecture is depicted in Fig. 2.

## 6 Algorithm for GAD

Algorithm: Genetic algorithm for determining the "i" reliable path.
buffer_size, gen, $B$m, $B$c, $a$0, the destination nodes $Y$, $X$.
1. Create the initial buffer_size
2. max $\leftarrow$ 1
3. While (max <= max) do
4. $B \leftarrow 1$
5. While ($B$ <= buffer_size) do
6. Getting chromosomes of the new buffer, select two chromosomes from the parent buffer based on $B$c. Apply crossover, and then alter the new child according to $B$m parameter.
7. Calculate the bandwidth of the new child (Band($B$))
8. If $X(B) \geqslant X$ then
Save this child as a contestant solution.
9. $B \leftarrow B + 1$
10. End if
11. End
12. Print all determined solutions
13. End

## 7 Genetic Algorithm

GA is an irregular pursuit strategy dependent on the development rule of science (natural selection), which was proposed by J. Holland in 1975 to streamline total NP issues [11]. The genetic algorithm imitates the cycle of Normal Choice (NS) hereditary calculation mirrors the cycle of normal choice (NS), which implies a focal idea of advancement. These attributes will be given to the future, which is additionally named as natural selection. GA can be viewed as a abstract representations of candidate solutions to resolve the optimization issues [12]. Hereditary calculations can be seen in the recreation where a populace of dynamic portrayals of up-and-comer arrangements to take care of an advancement issue [12]. GA solves a problem using an evolutionary approach by creating mutations to the current solution by selecting the better methods from this new generation and then using these improved methods to repeat the process. Since the genetic algorithm calculations are randomized hunt and improvement strategies, following the standards of development and common genetic qualities, have a lot of verifiable parallelisms. Hereditary calculations are randomized hunt and improvement strategies following the standards of development and common hereditary qualities, have a lot of verifiable parallelisms. Thusly, GA might be a piece of the more extensive class of organic cycle calculations that might be implemented to the streamlining of convoluted calculations, the instructing of text arrangement frameworks, and the development of smart counterfeit specialists that will decide the conditions randomly [13–17]. With these perceptions, genetic calculations perform a search in muddled, huge, partner proclaimed multimodal scenes and gracefully close ideal answers for target or wellness performance of an enhancement disadvantage. The essential module cycle of GA delineates in detail as follows [18]:

(i) Initially, build up a GA for a streamlining issue. So, the arrangement is a series of pieces that comprise a similar number of components.
(ii) An underlying populace is produced unexpectedly which ought to be spread over the inquiry space to speak to a wide assortment of arrangements.
(iii) The determination duplicates the string to be remembered for the future.
(iv) By choosing an irregular position, the hybrid is applied and it makes two new chromosomes.
(v) Determination and hybrid can create a surprising measure of contrasting strings.
(vi) The end model can be set by the number of development cycles, the measure of the variety of people of various ages.

## 8 Sybil Attack

Sybil attack is a genuine danger as it decreases the usefulness of VANET. Here, an attacker hub sends messages with various phony personalities to different hubs in the

organization. The attacker mimics a wide range of various hubs in the organization [19]. The hub forcing the characters of different hubs is called a malevolent hub, whose personalities are vanquished called Sybil nodes. Moderately, every other attack can be dispatched on an organization within the sight of Sybil attack. The chance could be thought of as a gridlock or mishap with the goal that different vehicles change their steering way or leave the street to serve the aggressor. Sybil attackers can likewise motivate false data in the organizations through some manufactured non-existing hubs [20]. For instance, on account of a mishap on an expressway, the main vehicle noticing the mishap is sending a change course/deceleration ready message to all the encompassing vehicles. The agents may advance this message to caution the supporters by assuming anything. This sending cycle can be disturbed by Sybil vehicles by not sending the admonition message. This may lead the travelers in the harmful way.

## 9 Result and Discussion

The source and destination are selected by the black dot that represents the vehicle. The lines joining them represent the path between source and destination. By clicking the go button on the right side, possible routes will be shown in Fig. 3.

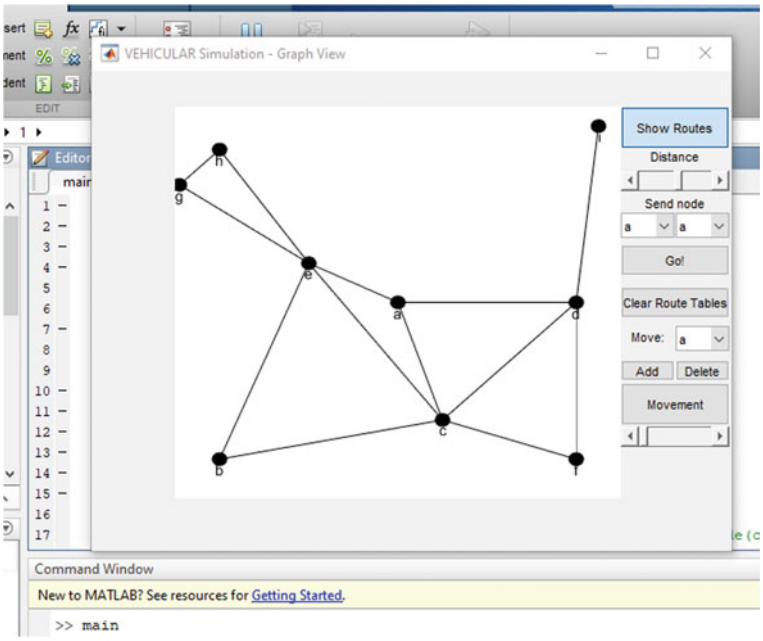



**Fig. 3** Initial configuration of VANET

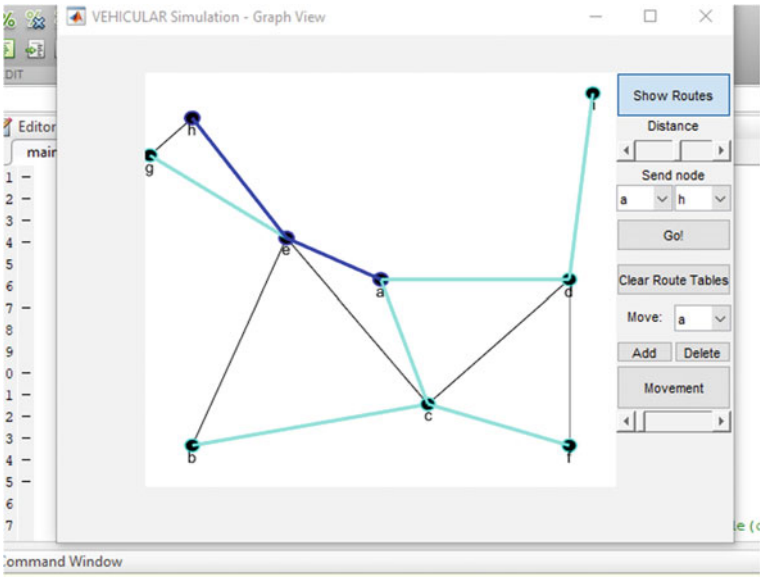



**Fig. 4** Identifying different routes for vehicles

From source to destination, the different routes are determined. The routes are established by using the AODV routing protocol. Initially, it will find all possible nodes between the source and the destination. Then, it selects the source and destination nodes. Both static and dynamic motions of vehicles are identified, and the path is established. From source to destination, different routes are determined and identify the reliable path using FCM. It is denoted in Fig. 4.

Using the DSDV algorithm, a table is created which holds the source node, destination node, packet sequence number, and hop distance. Identify the source and destination node to perform the desired communication. By clicking the movement button on the right side of the window, the nodes will dynamically change. It automatically chooses the shortest route identified by DSDV algorithm through which the message is to be transferred will be identified by using the DSDV algorithm. Finally, the table was automatically updated with the corresponding details, and it is represented in Fig. 5.

In Fig. 6, a total of 60 nodes, sink node, indicate as yellow color and it finds the cluster head, which is ready for attack. The cluster head node is indicated in white square color. The routing path is shown in magenta color. The routing path shows communication via source to destination. The test case performance is analyzed with different sensor radii like $R = [100, 120, 150, 170]$ m.

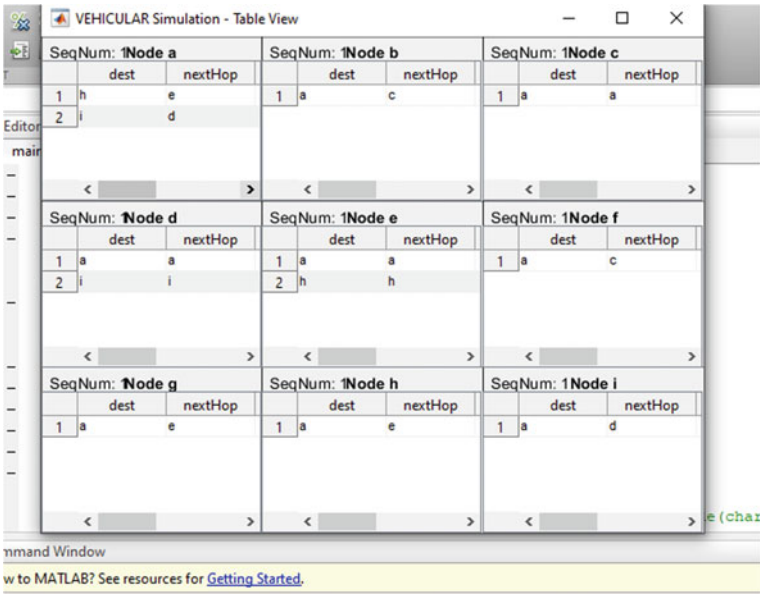



**Fig. 5** Determining hop count details

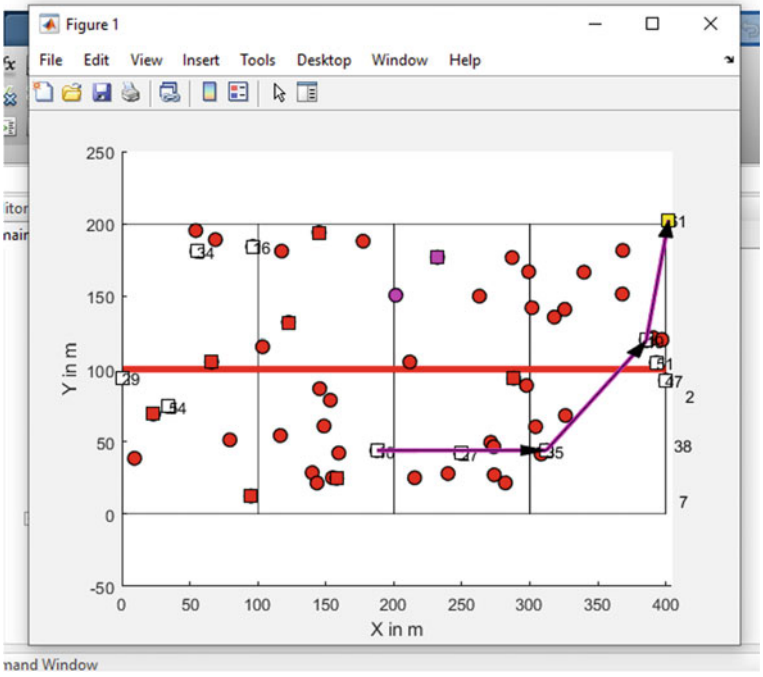



**Fig. 6** Clustering of nodes

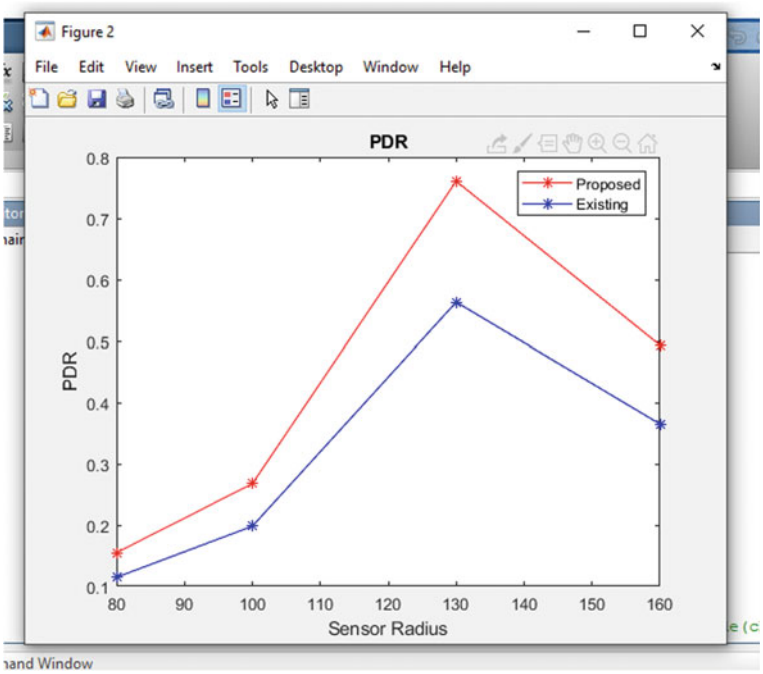



**Fig. 7** Packet delivery ratio

## *9.1 Packet Delivery Ratio*

The ratio of how many packets is successfully received by destination from the source node. A graph that compares the existing model with our proposed model is shown in Fig. 7. In this, the sensor radius is taken on the *X*-axis, and PDR is taken on the *Y*-axis. If the sensor radius is low, the communication path breakage may happen, so it automatically reduces the PDR; if the sensor radius increases more and most of the communication data from the source to sink with a low number of hopes in this time due to packet congestion, the PDR may decrease. But compared to the existing method, the proposed method has a high PDR.

## *9.2 Average End-to-End Delay*

The time taken for a communication across a network from source to destination is calculated difference b/w communication the data from source to the destination is calculated. The data transfer between source and destination may get lost due to node coverage area or their position, represented in Fig. 8.

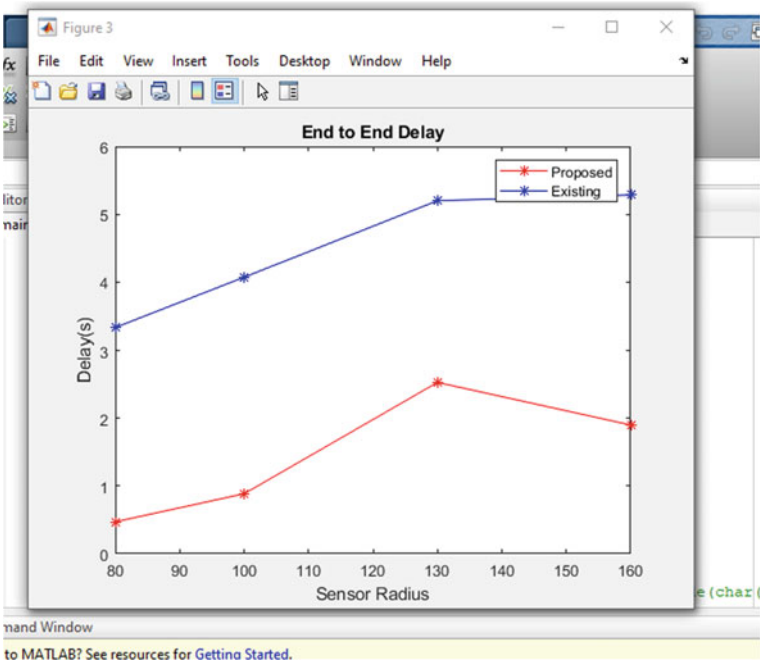



**Fig. 8** Average end-to-end delay

### *9.3 CH Formation Delay*

Initially, CH selection is based on a clustering algorithm, after the *t*-time slot CH update validates based on the neuro-fuzzy prediction. This time duration of CH update is called CH formation delay, represented in Fig. 9.

In Fig. 10, vehicles are moved in two lanes. Usually, the road side unit (RSU) is a wave device fixed along the road side for dedicated short-range communication. Every vehicle node communicates with the RSU. When the attacker spoof's multiple identities for one vehicle, 2 lanes mobility leads to Sybil attack to one vehicle, thus in 2 lanes mobility of vehicles leads to Sybil attack. Here, the attack is detected using the genetic algorithm, the node with the attack is cleared, and an encrypted message is safely sent to the required destination vehicle node.

## 10 Conclusion

In this article, the proposed algorithm called GAD is the combination of genetic, AODV, and DSDV algorithm for an effective vehicle-to-vehicle communication. The genetic algorithm determines the reliable route between the source and destination

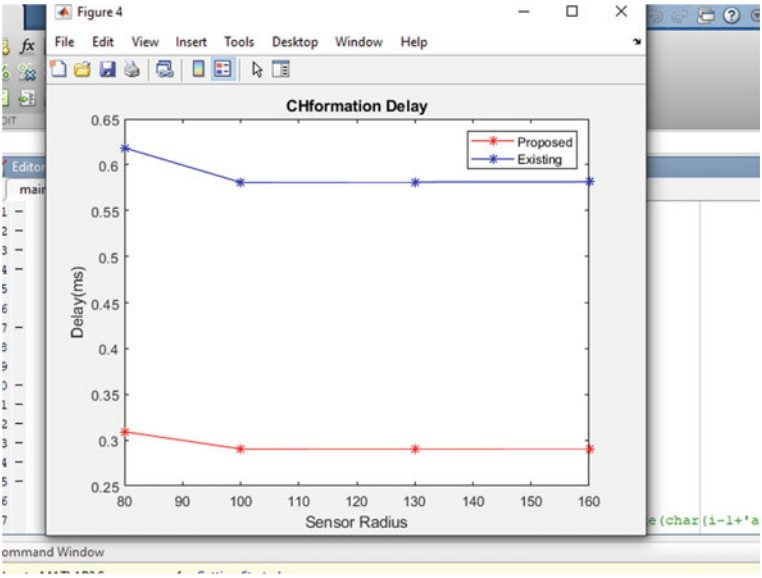



**Fig. 9** Channel selection

node routes by either AODV or DSDV. In the first part, the nodes are created, the communication links are established, and the shortest path between the nodes is found. In the second part, the messages are successfully sent to the receiver. And finally, in the third part, the Sybil attack was implemented and it was recovered using GAD protocol. The GAD protocol simulation results depicted that the efficiency is improved by following performance metrics, Packet Delivery Ratio (PDR), End-to-End delay and Cluster formation delay, and these metrics are compared with existing protocols AODV and DSDV represent that the efficiency is improved in the following performance metrics of PDR, End-to-End delays and cluster formation delay all are efficient compared to the existing method.

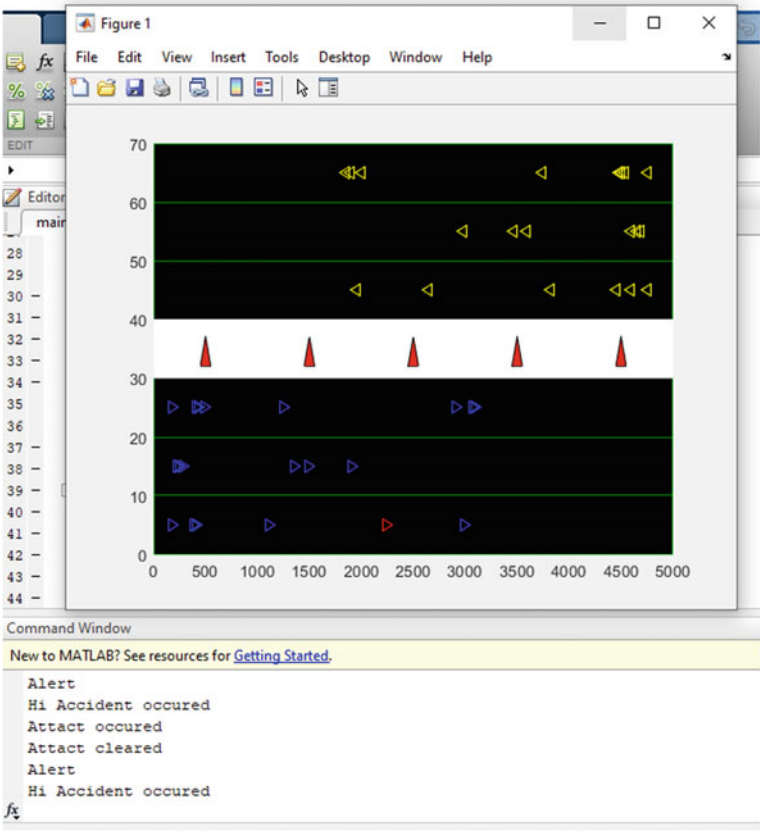



**Fig. 10** Effective communication using GAD protocol

## References


1. Alhan A et al (2015) Analysis of encryption Dgrp-data gather routing protocol based on Opnet in VANETs. IEEE. 978-1-5090-0076-0/15 $31.00 © 2015
2. Qian J, Jing T, Huo Y, Li Y, Zhou W, Li Z (2015) A next-hop selection scheme providing long path lifetime in VANETs. IEEE. 978-1-4673-6782-0/15/$31.00 ©2015
3. Martin IV, Urquiza-Aguiar L, Aguilar-Igartua M, Guerin-Lassous I Transient analysis of idle time in VANETs using Markov-reward models. IEEE. https://doi.org/10.1109/tvt.2017.2766449
4. Kochhar R Performance study of VANET using ant based routing algorithms. IEEE. 978-9-3805-4416-8/15/$31.00 c 2015
5. He J, Cai L, Pan J, Cheng P Delay analysis and routing for two-dimensional VANETs using carry-and-forward mechanism. IEEE. https://doi.org/10.1109/tmc.2016.2607748
6. Wang H, Liu RP, Ni W et al (2015) VANET modeling and clustering design under practical traffic, channel and mobility conditions. IEEE Trans Commun 63(3):870–881
7. Yang F, Tang Y (2014) Cooperative clustering-based medium access control for broadcasting in vehicular ad-hoc networks. IET Commun 8(17):3136–3144

8. Yang F, Tang Y, Huang L (2014) A multi-channel cooperative clustering-based MAC protocol for VANETs. In: Proceedings of the IEEE wireless telecommunications symposium (WTS), Washington, USA, pp 1–5
9. Rajalakshmi D, Meena K (2020) An efficient selfishness control mechanism for mobile Ad hoc networks. Int J Adv Res Eng Technol 11(7)
10. Su H, Zhang X (2007) Clustering-based multichannel MAC protocols for QoS provisioning over vehicular ad hoc networks. IEEE Trans Veh Technol 56:3309–3323
11. Ren M, Zhang J, Khoukhi L, Labiod H, Vèque V (2018) A unified framework of clustering approach in vehicular ad hoc networks. IEEE Trans Intell Trans Syst 19(5):1401–1414
12. Hartmann T, Kappes A, Wagner D (2016) Clustering evolving networks. In: Algorithm engineering. Springer, Cham, Switzerland, pp 280–329
13. Dror E, Avin C, Lotker Z (2011) Fast randomized algorithm for hierarchical clustering in vehicular ad-hoc networks. Ad Hoc Networks 11(7)
14. Eiza MH, Ni Q (2013) An evolving graph-based reliable routing scheme for VANETs. IEEE Trans Veh Technol 62(4):1493–1504
15. Rajalakshmi D, Meena K (2020) A hybrid intrusion detection system for mobile Ad hoc networks using FBID protocol. Scalable Comput Pract Experience 21(1)
16. Mitra S, Jana B, Poray J A novel scheme to detect and remove black hole attack in cognitive radio vehicular ad hoc networks (CR-VANETs). https://doi.org/10.1109/iccece.2016.8009589
17. Dilli R (2018) Performance analysis of look up protocol for VANET information retrieval services. In: 9th ICCCNT
18. Gao N, Tang L, Li S et al (2014) A hybrid clustering-based MAC protocol for vehicular ad hoc networks. In: Proceedings of the IEEE international workshop on high mobility wireless communications (HMWC), Beijing, China, pp 183–187
19. Hafeez KA, Zhao L, Liao Z, Ma BN-W (2012) A fuzzy-logic-based cluster head selection algorithm in VANETs. Proc IEEE Int Conf Commun 1(1):203–207
20. Li W, Tizghadam A, Leon-Garcia A (2012) Robust clustering for connected vehicles using local network criticality. In: Proceedings of the IEEE international conference on communications, vol 1, no 1, pp 7157–7161

# Performance Study of Free Space Optical System Under Varied Atmospheric Conditions


**Hassan I. Abdow and Anup K. Mandpura**

**Abstract** In this paper, the performance study of free space optical (FSO) systems is completely analyzed, when the atmospheric channel is affected by atmospheric conditions such as haze and fog. Comparison of the FSO system's performance with semiconductor optical amplifier (SOA) and erbium doped fiber amplifier (EDFA) under the influence of haze and fog is studied. The received power, eye diagram, and quality factor (Q-factor) are the performance metrics taken into consideration in this paper. Through simulations, it is demonstrated that the fog has a detrimental effect on FSO system performance, when compared with the haze. Further, the distance over which the FSO system can work reliably will be improved by using EDFA in the place of SOA for performing pre-amplification and post-amplification.

**Keywords** Free space optical (FSO) · Atmospheric channel · Erbium doped fiber amplifiers (EDFA) · Quality (Q) factor · Semiconductor optical amplifiers (SOA) · Fog and haze

## 1 Introduction

In the past few decades, there has been an ever-increasing demand for high bandwidth along with the popularity of live streaming applications for work, social interaction, entertainment, and infotainment purposes. Various research studies will leverage research significance to improve the spectral efficiency of the existing cellular system and accommodate these demands [1–5]. The strategies in [1–5] are promising; however, an alternative solution for decongesting the cellular spectrum is to shift from the existing radio frequency carrier range of 300 MHz to 3 GHz to optical frequency range of 300–3000 GHz. The main advantage being that, licensing of spectrum in the optical domain is not required. Therefore, the optical domain offers a large bandwidth that is economical and helps increase the capacity of the communication system [6]. Recently, free space optical (FSO) communication has been widely studied in

H. I. Abdow (✉) · A. K. Mandpura
Departmentof Electrical Engineering, Delhi Technological University, Delhi, India
e-mail: hassan@uonbi.ac.ke



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_64

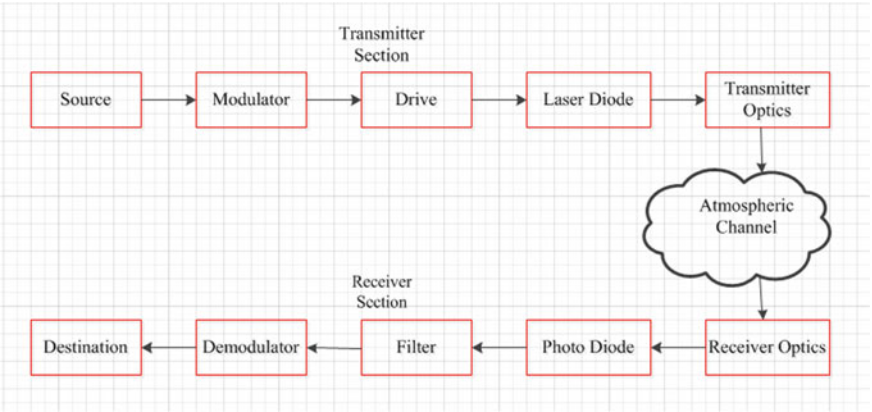



**Fig. 1** FSO communication system

[7–11] due to the dual advantages of optical spectrum (large bandwidth), licensed free operation, security, ease of implementation, absence of electromagnetic interference, and wireless transmission channel. Due to the scarcity of spectrum for radio frequency (RF) transmission, these features offer an important solution for wireless access [12]. FSO is communication technology where information is exchanged by sending a beam of light from the transmitter to the receiver. Figure 1 block diagram of the FSO communication system is presented. The information is transmitted from the source to the receiver through the atmospheric channel. The source data is sent to the modulator, and the modulated carrier signal, with the information, is transmitted through the FSO channel. The information at the receiver is sent to the photodetector, filter, and demodulator for decoding at the destination [13]. This technology finds applications where the laying of physical cable is impossible due to the terrain of the area, for example, in a hilly location. The transmission of information in FSO is similar to that in optical fiber communication, i.e., the transmitted data is modulated by a laser light. However, transmission medium in FSO is the atmosphere, unlike the optical fiber where the light travels through a fiber. FSO communication system performance is adversely affected by the environment through which it propagates. The fluctuations of pressure and temperature variations lead to the refractive index which results in atmospheric turbulence. This atmospheric turbulence leads to an increase in the systems errors, therefore, degrading the FSO system performance [14].

The FSO communication system performance largely depends on weather factors such as fog, rain, and haze [15]. These atmospheric disturbances attenuate and scatter the light beam, thereby hindering the direct path of the laser light from the transmitter to the receiver [16]. These disturbances therefore significantly affect the performance of FSO communication, thus, reducing the overall range and capacity of the system. It is therefore of interest to study the performance of FSO system when the transmission channel is affected by atmospheric disturbances such as fog and haze. Free

space optical communication link performance is also degraded due to turbulence in atmospheric conditions and other parameters such as distance and optical power. Due to the loss in signal strength caused by atmospheric disturbances and obstacles, it is difficult to attain a high performance in a free space optical network [17]. In real-life scenarios, the availability of FSO communication link is limited during fog and haze. The intensity of the FSO signal is subjected to random fluctuation due to atmospheric turbulence. Scintillation causes performance degradation and potential loss of signal connectivity. These shortcomings pose the main challenge to FSO communications system deployment. Therefore, different channel modeling and diversity techniques have been studied to improve system performance [18]. The quality factor, received power, and span of the system (distance) are the performance metrics for a FSO system.

In this work, FSO communication system's performance under atmospheric conditions such as fog and haze are studied. FSO suffers from signal degradation due to external and internal impairments, therefore, to compensate for the attenuation caused by atmospheric conditions; pre-amplification and post-amplification of the signal are performed. Before the advent of optical amplifiers, the signal amplification was performed by optical to electrical and then to optical conversion for a single wavelength and moderate data rate requirements. However, this amplification process is not preferable for high data rate and multiple wavelengths [19]. Moreover, this technique increases equipment cost and introduces delay and noise due to large numbers of regenerators for long haul transmission. With the introduction of optical amplifiers, the delay and congestion issue can be overcome by simultaneous amplification of various wavelengths [20]. The repeaters have been replaced with amplifiers, in order to have a cheaper alternative for optical amplification and transceivers in [21]. Erbium doped fiber amplifiers (EDFA) and semiconductor optical amplifiers (SOA) are the most commonly used optical amplifiers. Among these amplifiers, EDFA is conventionally used for long haul optical transmission over fiber. EDFA's main benefits over other amplifiers are high-power efficiency, low-noise figure, and commercial availability in the C-band and L-band [22]. SOA is used as a preamplifier before the signal reaches the receiver to boost the signal level. One of the feature of SOA is low-noise value of 5–8 dB as compared to EDFA of 4 dB. However, it is hard to obtain power levels above 10 mW due to the relatively small output saturation power values of around 5 mW. Parameters such as polarization, nonlinear effects, sensitivity, and high losses at the junction make these amplifiers more challenging to use than the inline amplifiers. EDFA has proven to be best for signal amplification [23]. EDFA can amplify different channels without reducing its bandwidth. The difference between EDFA and SOA is an active region where the gain generation occurs. EDFA generates directly within the glass fiber, while SOA is directly within the structure of the semiconductor [24].

There have been studies on multiple transceivers and channel modeling to reduce the effect of atmospheric conditions. Diversity techniques are used to transmit the information and enhance system efficiency [25]. In [26], the FSO communication link performance is analyzed under foggy weather conditions employing different transmitters/receivers pairs and an EDFA amplifier using OptiSystem software for

simulation. Free space optical link analysis is carried out and evaluated based on Q-factor for various modulation techniques and weather conditions. In [27, 28], by using log-normal and Gamma–Gamma scintillation models, the authors analyzed the value of Q-factor versus proposed distance of FSO link for varying range of index of refraction. Therefore, in this work, the objective is to study the FSO system performance with pre-amplification and post-amplification using EDFA. The system performance is evaluated using a received power, eye diagram, and Q-factor.

The paper is organized as follows: Section 1 discusses the introduction to FSO and the performance metrics used in the paper. Section 1.1, discusses the FSO system model with EDFA at both ends. Section 2 presents results and compares an FSO system's performance with EDFA and SOA. Section 3 provides conclusions.

## *1.1 The System Model*

Figure 2 shows a system model of free space optical communication. It comprises of a transmitter, atmospheric channel, and a receiver. The transmitter comprises of a light source, modulator, data source generator, and an amplifier. The main function of the transmitter in FSO is to convert the data into an optical signal that can propagate through the atmosphere to the receiver. The characteristics of a transmitted signal are affected in FSO channel due to the various weather conditions. The receiver comprises of an amplifier, photodetector, low-pass filter, and an analyzer. The optical signal at the receiver is amplified by the amplifier, then the photodetector is used to convert the received optical signal into an electrical signal, and then the low-pass filter is used to remove the unwanted high-frequency noise signal.

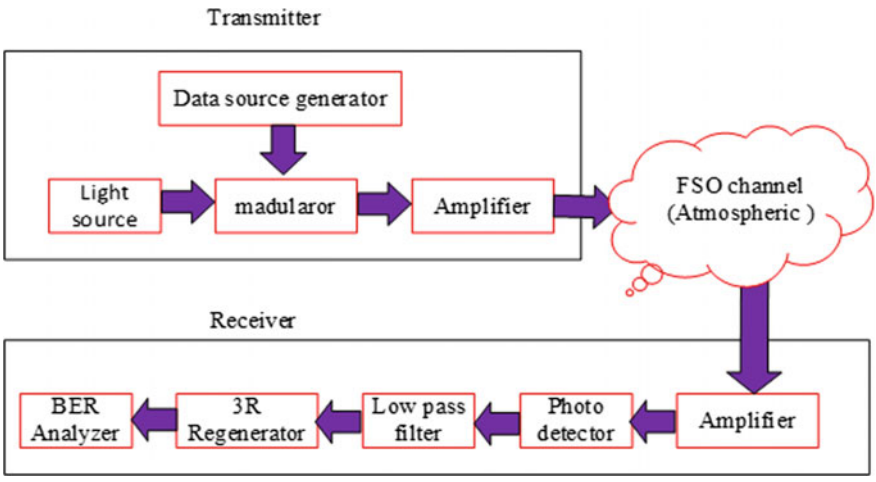



**Fig. 2** FSO system model

**Table 1** Parameters of FSO used for the simulations

| Parameter | Values |
|---|---|
| Data rate<br>Modulation format | **EDFA** 40 Gbps, **SOA**<br>0.5 Gbps<br>Non-return to zero (NRZ) |
| CW (continuous wave) laser source | 25 dBm |
| Aperture diameter transmitter | 2.5 cm |
| Aperture diameter receiver | 20 cm |
| Operating frequency | 193.414 THz |
| Responsivity | 1 A/W |
| Attenuation-very clear | 0.1 dB/km |
| Attenuation-haze | 4.2 dB/km |
| Attenuation-light-fog | 15.5 dB/km |
| **SOA-parameters**<br>Injection current | 0.087 A |
| Width | 3e−006 |
| Height | 6e−007 |
| Confinement factor | 0.15 |
| Enhancement | 5 |

This paper studies the FSO communication system's performance for different weather conditions as follows: very clear weather, haze, and fog. The performance metrics used in the study are Q-factor, received power, and eye diagram for the above different weather conditions. The FSO system simulation is modeled using OptiSystem software. The system parameters and different values of attenuation of each of the atmospheric conditions are listed in Table 1 [29, 30].

## 2 Result and Discussion

This section discusses the impact of using EDFA as a preamplifier and post-amplifier in free-space optical system for different atmospheric conditions. The free space optical system was simulated using OptiSystem as shown in Fig. 3 and analyzed under different atmospheric conditions. The system parameters are given in Table 1. The pseudorandom binary sequence generator is used to generate the data bits, which are fed to Mach–Zehnder modulator for conversion to an optical signal. A light source of wavelength 1550 nm transmits power of 25 dBm. The modulator's output is fed to an optical preamplifier (EDFA or SOA) and transmitted through the FSO channel as shown in Fig. 3. The received optical signal from the receiver is converted into an electrical signal using a photodetector. The low filter rejects the high-frequency noise.

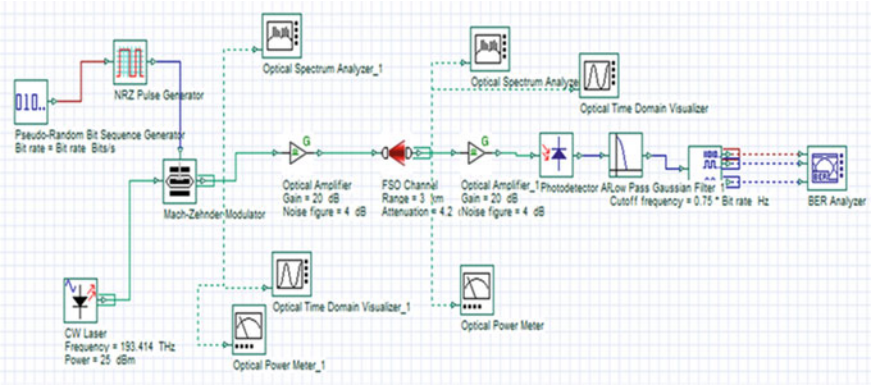



**Fig. 3** Model of a FSO system in OptiSystem

Figure 4 compares Q-factor versus distance of FSO system with EDFA and SOA for different weather conditions, i.e., very clear weather, haze, light fog, and moderate fog. It can be observed from the figure that FSO with EDFA has a better Q-factor over a large distance in comparison to SOA. This is due to the fact that the noise figure of SOA is greater than 5 dB whereas the noise figure of EDFA is less than 5 dB [31]. The typical gain of EDFA is also greater than 40 dB, while that of SOA is greater than 30 dB [32, 33]. The FSO system with EDFA has a better performance in comparison to SOA for all different weather conditions. As the distance is increasing, the quality factor is decreasing for both amplifiers. Very clear weather has the least

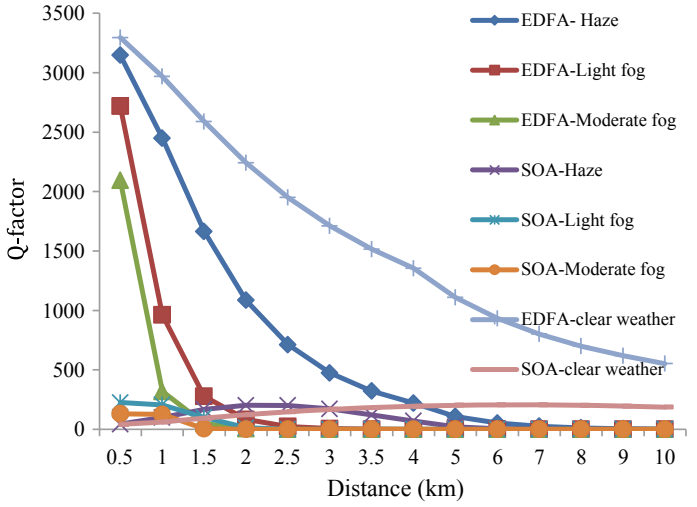



**Fig. 4** Q-factor versus distance for different atmospheric conditions (haze, light fog, moderate fog and very clear weather)

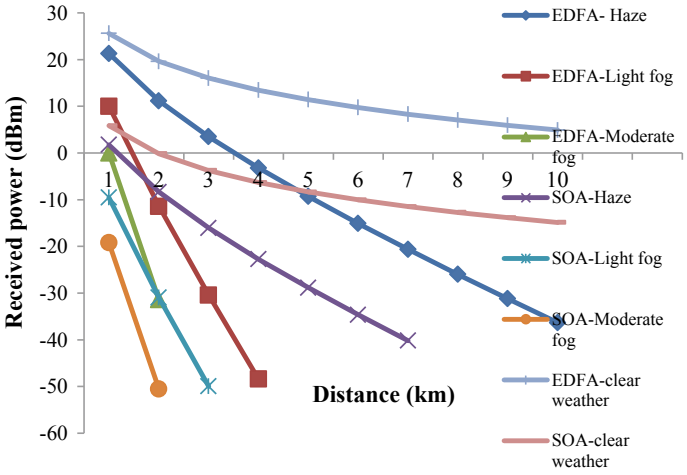



**Fig. 5** Received power versus distance for different atmospheric conditions (light fog, haze, moderate fog, and very clear weather)

effect on the performance of FSO as compared to haze and fog. Moderate fog is the worst among all different weather conditions for FSO system.

Figure 5 shows the relationship between the received power and transmission distance for different weather conditions by using EDFA and SOA. Comparing between the two amplifiers, EDFA has a better received signal power than SOA under the similar weather conditions. It can be observed from the figure that the received power of an FSO system with EDFA is larger in comparison to an FSO system with SOA for a distance of 10 km for very clear weather conditions, and the received power with EDFA is 4.692 dB, whereas the received power with SOA is −14.852 dB. From the figure, it is evident that the received signal power decreases with increasing distance. The gap between the received powers of FSO with EDFA and FSO with SOA is nearly 19.544 dB for a distance of 1–10 km for haze weather conditions. The difference in received power of FSO with EDFA and FSO with SOA for haze is approximately 19.544 dB for communication over short distances. Again, it is observed that very clear weather conditions have the least affected on FSO system performance in comparison to haze and fog for all the cases.

Figures 6, 7, and 8 show eye diagrams of FSO link for haze at 9 km, moderate fog at 2 km, and clear weather at 70 km, respectively. The eye opening signifies the performance of an FSO system; if the opening is wide, the system performance is good, and if the opening is narrow, the performance is poor. The qualitative analysis of digital signal transmission can be studied using the eye diagram. In Fig. 8, the system has a large eye opening between the upper and lower levels as compared to Figs. 6 and 7. The height of the eye opening determines the noise margin of the system at a specified time interval. It can be concluded from the three figures that very clear weather conditions have wider opening and hence a better performance than the fog and haze weather conditions. The simulation results show that both

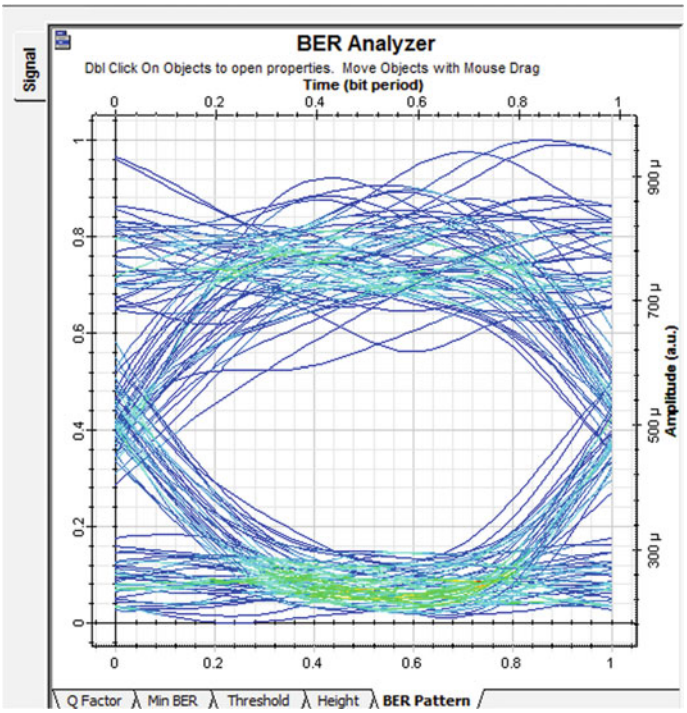



**Fig. 6** Eye diagram for haze at 9 km

distance and attenuation affects the FSO link. The opening of the eye for haze at 9 km is 2.3e−04 while it is 2.769e−04 for moderate fog at 2 km.

## 3 Conclusion

In this paper, the performance of free-space optical (FSO) communication system with EDFA in the presence of haze and fog is studied. The FSO system performance was compared using received power, Q-factor, and eye diagram for different weather conditions. From the comparison of the results, it is concluded that fog has a detrimental effect on FSO system performance in comparison to haze.

Under clear weather conditions, the FSO system with EDFA has a range of 67.1 km, whereas for haze and moderate fog conditions, the range is 9 km and 2.5 km, respectively. Also, from the results presented, it can be summarized that FSO system performance is enhanced by using EDFA amplifier in comparison to a SOA amplifier.

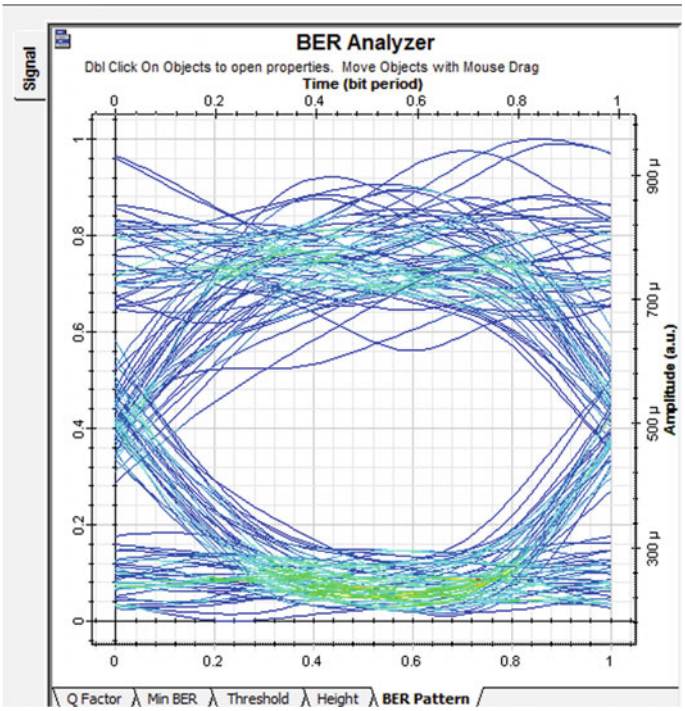



**Fig. 7** Eye diagram for moderate fog at 2.5 km

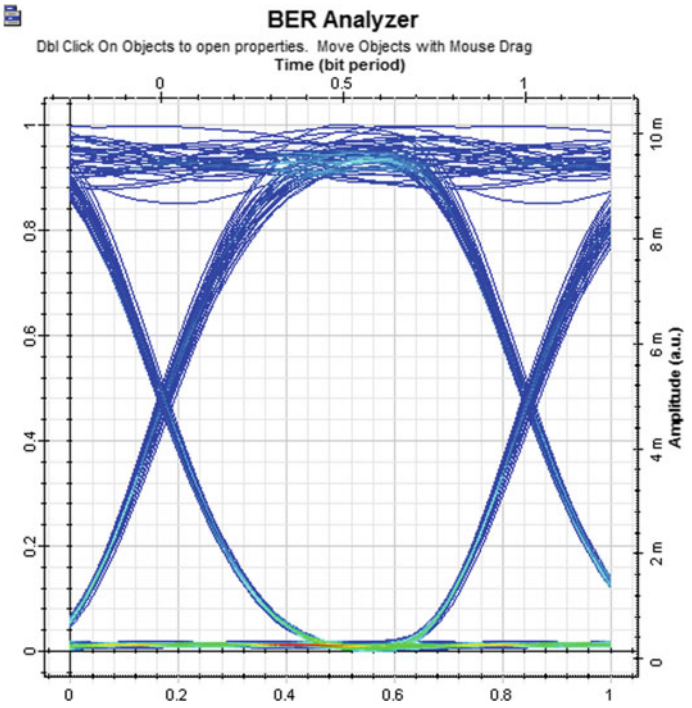



**Fig. 8** Eye diagram at 70 km for very clear weather

**Acknowledgements** I would like to express my sincere appreciation and gratitude to Indian Council for Cultural Relations, Govt. of India for enabling me to pursue my Ph.D. research. Special thanks also go to the faculty in the Department of Electrical Engineering for their advice and providing an enabling environment. I also express my gratitude to my friends for their valuable time.

## References

1. Ahmed A, Singh A, Singh A, Kaur S (2019) Performance analysis of WDM-MIMO free space optical system under atmospheric turbulence. 978-1-7281-1380-7. IEEE
2. Malhotra R, Miglani JS (2019) Performance enhancement of high capacity coherent DWDM free space optical communication link using digital signal processing. Springer Nature
3. Hong Y-Q, Shin W-H, Kwon D-H, Han S-K (2019) High PDG- OA-based MLPolSK modulation for spectral efficient free space optical communication. IEEE 32(1)
4. Aarthi G, Ramachandra Reddy G (2018) Average spectral efficiency analysis of FSO links over turbulence channel with adaptive transmissions and aperture averaging (Elsevier)
5. Al-Nahhal M, Ismail T (2019) Enhancing spectral efficiency of FSO system using adaptive SIM/M-PSK and SIMO in the presence of atmospheric turbulence and pointing errors. John Wiley & Sons Ltd

6. Tabassum N, Franklin N, Arora D, Kaur S (2018) Performance analysis of free space optics link for different cloud conditions. In: IEEE, 2018, 4th international conference on computing, communication and automationional (ICCCA)
7. Anbarasi K, Hemanth C, Sangeetha RG (2017) A review on channel models in free space optical communication systems. Elsevier, Amsterdam
8. Sawhil, Agarwal S, Singhal Y, Bhardwaj P (2018) An overview of free space optical communication. Int J Eng Trends Technol (IJETT)
9. Raj AB, Majumder AK (2019) Historical perspective of free space optical communications: from the early dates to today's developments IET
10. Gupta D, Sharma P, Tandon R, Sharma H, Gupta M (2018) Free space optical communication. Int J Sci Tech Adv 4(1):55–60. ISSN: 2454-1532
11. Trichili A, Cox MA, Ooi BS, Alouini M-S (2020) Roadmap to free space optics. J Opt Soc Am B (JOSA B)
12. Mazin AAA (2016) Performance analysis of terrestrial WDM-FSO link under different weather channel. World Scientific News 33–44
13. Mahajan S, Prakesh D, Singh H (2019) Performance analysis of free space optical system under different weather conditions. In: 2019 6th International conference on signal processing and integrated networks (SPIN). IEEE, pp 220–224
14. Sawhil, Agarwal S, Singhal Y, Bhardwaj P (2018) An overview of free space optical communication. Int J Eng Trends Technol (IJETT) 120–125
15. Yadav SK, Chouhan S (2014) Performance analysis of optical wireless communication link by multiple Tx/Rx with and without amplifier. Int J Eng Res Technol (IJERT) 3(6)
16. Alkholdi AG, Altowij KS (2014) Climate effects on performance of free space optical communication systems in Yemen. Optoelectron 7(1):91–101 (Higher Education Press and Springer-Verlag, Berlin Heidelberg)
17. Anis AA, Rashidi CBM, Aljunid SA, Rahman AK (2018) Evaluation of FSO system availability in haze condition. IOP Conf Ser Mater Sci Eng 318:012077 (IOP Publishing). https://doi.org/10.1088/1757-899x/318/1/012077
18. Khan MN (2014) Importance of noise models in FSO communications. J Wirel Commun Networking 10
19. Dutta MK (2017) Design and performance analysis of EDFA and SOA for optical WDM networks: a comparative study (IEEE)
20. Singh S, Singh A, Kaler RS (2011) Performance evaluation of EDFA, RAMAN and SOA optical amplifier for WDM systems
21. Shakya S (2019) Machine learning based nonlinearity determination for optical fiber communication-review. J Ubiquitous Comput Commun Technol (UCCT) 121–127
22. Abubaker A, Ibrahim NM (2014) Comparison of optical amplifiers in optical communication systems EDFA, SOA and Raman 09:8738–8741 Int J Current Res 6
23. Ivaniga T, Ivaniga P (2017) Comparison of the optical amplifiers EDFA and SOA based on the BER and *Q*-Factor in C-Band. Adv Opt Technol 1–9
24. Gupta U, Rani M, Goyal R (2016) Comparison of different amplifiers at different data rates in WDM system performance. Int J Res Educ Sci Methods (IJARESM) 98–101
25. Al-Juboori S, Fernando X (2019) Characterizing a decorrelator for selection combining receivers in Nakagami-m fading channels. Int J Electron Commun (AEÜ) 19
26. Singh M (2016) Mitigating the effects of fog attenuation in FSO communication link using multiple transceivers and EDFA. J Opt Commun 6
27. Kaur S, Kakati A (2018) Analysis of free space optics link performance analysis of free space optics link performance and modulation formats for terrestrial communication. J Opt Commun 6
28. Singh H, Chechi DP (2019) Performance evaluation of free space optical (FSO) communication link: effects of rain, snow and fog. In: 2019 6th International conference on signal processing and integrated networks (SPIN), 387–390
29. Kaur H, Sarangal H (2015) Impact of various weather conditions on free space optics using 4X4 transmitter/receivercombination integrated with different ways of amplification. Int J Adv Res Comput Commun Eng 388–393

30. Sharma A, Utreja B (2019) Performance collation of different spectrum slicing techniques in FSO systems. Compliance Eng J 246–257
31. Sharma C, Singh S, Sharma B (2013) Investigations on bit error rate performance of DWDM free space optics system using semiconductor optical amplifier in intersatellite communication. Int J Eng Res Technol 2(8)
32. Mohammed KA, Younis BMK (2020) Comparative performance of optical amplifiers: Raman and EDFA. Telkomnika Telecommun Comput Electron Control 18(4). ISSN: 1693-6930
33. Kapse MC, Shriramwar SS (2017) Performance of various types of amplifiers in DWDM technology. Int J Latest Trans Eng Sci 1(3). ISSN: 2321-0605

# Malicious URL Detection Using Machine Learning and Ensemble Modeling


**Piyusha Sanjay Pakhare, Shoba Krishnan, and Nadir N. Charniya**

**Abstract** Websites are software applications that allow us to connect and interact with the data located in the web servers. Websites allow the user to capture, store, process, and exchange sensitive data like banking details and personal details. Web pages are accessed by merely entering the required URL in the browser. To prevent sensitive information from users, the attackers/hackers make duplicate websites and send them to victims through phishing emails. In this article, the machine learning framework is used to find malicious URLs. Here, five different machine learning algorithms such as the logistic regression algorithm, K-nearest neighbor algorithm, decision tree algorithm, random forest algorithm, and support vector machine algorithm have been used. An ensemble modeling has been done using these algorithms, and the performance of each algorithm has been compared.

**Keywords** Cyberattacks · Malicious URLs · Supervised learning · Machine learning · Ensemble models

## 1 Introduction

The epoch of big data where data in quintillions is generated every single day [1]. With more and more data on the Internet that included user's data, credit card details, phone numbers, etc., hackers are trying to obtain those data by stealing it from the user unknowingly. This is done by sending various phishing e-mails, creating duplicate websites, sending attachments via e-mail, and SMSs. Phishing uses e-mails as a weapon. The attacker dissembles as a trustworthy person or organization and sends a

P. S. Pakhare (✉) · S. Krishnan · N. N. Charniya
V.E.S Institute of Technology, Mumbai, India
e-mail: 2018.piyusha.pakhare@ves.ac.in

S. Krishnan
e-mail: shoba.krishnan@ves.ac.in

N. N. Charniya
e-mail: nadir.charniya@ves.ac.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_65

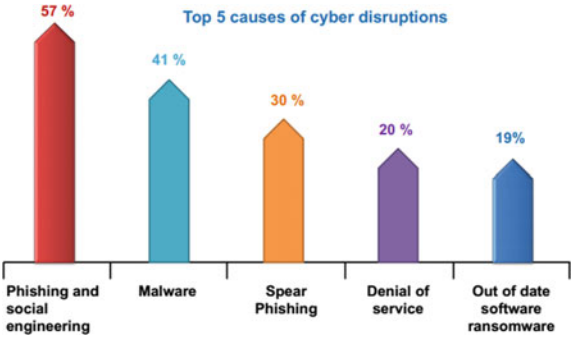



**Fig. 1** Causes for cyber disruptions [3]

message to be from the bank, asking for details or some note from companies tricking the recipient to open attachment or website links. The spreading of coronavirus across the world has caused people to work remotely from their homes. URL is the entry point of most cyber-attacks. If the attacker has control of the malware-hosting server, he/she can easily change the URL address making it point toward the malicious content [2]. As a result, it will risk the entire organization's privacy. URLs are classified into two (binary) categories—malicious and legitimate.

There are various domains on the Internet that have words like "covid19", "coronavirus", etc., cybercriminals are creating hundreds or even thousands of new websites every day to carry out phishing, fraud, and malware. The percentages of major cyber-attacks are shown in Fig. 1.

From Fig. 1, it is evident that phishing and malware attacks are the most common attacks and these attacks are spread using links and e-mails. Apart from these, URLs are host to cross-site scripting (XSS) attacks and URL injection attacks. According to Data Breach Investigating Report (DBIR), 32% of the confirmed data breaches involve phishing. Approximately 80% of the email attacks are malware-less, the attackers use spear-phishing and impersonation tactics. Wandera's 2020 Mobile Threat Landscape Report showed that more than 50% of all surveyed organizations have experienced a minimum of one mobile phishing incident [4]. Barracuda researchers [5] detected 467,825 spear-phishing attacks through email between March 1, 2020, to March 23, 2020, out of which 9116 were related to coronavirus.

For detecting these malicious attacks, machine learning (ML) is the most popular approach used. ML is a partial application of artificial intelligence. It is used to develop algorithms that can access the data and can use it for self-learning [6]. These algorithms work efficiently when huge data is involved. ML is often categorized into three, viz. supervised learning, unsupervised learning, and reinforcement learning [7]. All the ML algorithms are labeled training data with specific inputs, outputs, and system parameters. The proposed system uses supervised learning algorithms where past data is utilized to train the model to predict future outcomes. In unsupervised learning, the data used to train the model is neither labeled nor classified. Apart from these two categories, there is a semi-supervised learning algorithm that uses both unlabeled and labeled data for training. There are various ML algorithms, viz.

Regression, Classification, Clustering, Recommendation System, etc. Classifying the malicious URLs requires classification algorithms mentioned below.

This article deals with detecting the malicious URLs using various machine learning algorithms, namely [8],

(a) Logistic regression-It gives a probability as if a particular event will occur or not using a logistic function based on the input.
(b) Decision tree-The data is split at various nodes and the end decision (good/bad) is given by the leaves of the tree.
(c) Random forest-It is a group of decision trees and the final output is selected by majority voting.
(d) K-nearest neighbors-This algorithm will classify the sample point based on the classification of its k-neighbors. The best way to find the optimum value of k is using the elbow method. Usually, k's value should be odd for a classification problem.
(e) Support vector machine (linear, RBF, and sigmoid kernels)-for classification, a hyperplane/decision boundary is defined which separates the data and classifies them into two categories—good (legitimate) and bad (malicious). If the problem cannot be separated linearly, then kernels are used. Kernels are complex mathematical functions.

Further, an ensemble method is used to improve the accuracy of the algorithms. Ensemble modeling is the process where various algorithms are used to predict an outcome. It is a combination of individual classifiers that helps in the classification of new test samples. Ensemble learning in supervised learning has become one of the in-depth areas of study among machine learning researchers [9]. The majority voting method of ensemble models is utilized. Finally, these algorithms were tested completely on a new data to find which algorithm gives better classification. The motivation for using ensemble modelling is that it reduces the generalization error. Even though it consists of multiple algorithms, it acts as a single model.

This article is organized as follows: In Sect. 2, some literature review is done. In Sect. 3, our approach, methodology, and flowcharts used for URL classification are shown. In Sect. 4, the results of the experiments are shown and finally, in Sect. 5, the conclusion for the research work is presented.

## 2 Literature Review

Various studies have been conducted to find a solution to this problem using different techniques.

Vanhoenshoven et al. [10] have used a public dataset having 2.4 million URL instances to classify good and bad URLs. Having around 3 million features, they divided the features into 3 sets. Set A contains real and binary values, Pearson's correlation was found and values having more than 0.2 coefficient value were selected. Set B has only binary values and attributes having more than 0.1 coefficient are selected.

Set c has only real-valued data which are binary having non-zero values. They have classified the performance of the algorithms, viz. Naive Bayes, k-nearest neighbors, support vector machines, multi-layer perceptron, decision trees, and random forest. The algorithms were evaluated based on accuracy, precision, and recall. The random forest classification method showed maximum accuracy, recall, and precision.

Tan et al. [11] have addressed the problem of concept drift. Concept drift is the phenomenon where the distribution of data changes, i.e., malicious URL may turn into benign URL. If it happens, then the trained model cannot accurately tell the classification of such a URL. Three features have been selected which are IP blocks, destination port, and domain. Their system framework consists of four modules—Feature Extraction, Training, Prediction ad Concept Drift Detection. The functioning of the concept drift algorithm is evaluated by comparing the WRST algorithm with the CUSUM algorithm. The WRST algorithm had the highest accuracy and minimum delay over the CUSUM algorithm.

Feroz and Mengel [12] have developed a method that will automatically classify the URLs based on their host-based and lexical features. They have combined clustering and classification method. The results are divided into three categories—severe, moderate, and benign. Additionally, an internal scale is given to each classification—Severe—red, moderate—yellow, benign—green. Their classification accuracy is 98.46%.

Manjeri et al. [13] have experimented on classifying the URLs for two cases—class imbalance and without class imbalance. Recursive Feature Elimination technique is used to narrow down to eight features for classification and further ranking them according to their priority. Random forest algorithm gave the maximum accuracy of 96%. Finally, association rule mining was performed to find the relation between variables in data using three algorithms—Apriori, FP Growth, and decision tree.

## 3 Experiment

### *3.1 Url Structure*

Figure 2 shows the structure of a URL.


**Fig. 2** Structure of URL [14]

1. Protocol/Scheme

   This indicates the webserver which type of server to utilize when it accesses the page on the website. Nowadays, Hypertext Transfer Protocol Secure i.e., HTTPS is the commonly used protocol.

2. Hostname

   Hostname consists of Top Level Domain (.com, .net, .org, .uk), Domain Name (google.com, urlinfo.com) and Subdomain (www)

3. Port

   Port number is used to access a server application that is running on the machine. By default, HTTP uses port 80 and HTTPS port 443.

4. Path

   The path describes the specific source in the host that is accessed by the web client

Hackers/Cybercriminals will change certain parts of the URL to carry out attacks like phishing.

### *3.2 Dataset*

The dataset is obtained from GitHub [15]. It contains 344,821 good URLs and 75,643 bad URLs. To avoid imbalance classification problems, the dataset ia made even by eliminating random good URLs. So, the new dataset has 75,643 good URLs and 75,643 bad URLs.

### *3.3 Pre-processing*

As the dataset contains text data, it is first pre-processed by using tokenization, where the URLs are divided into tokens. Since the data is a URL, the URLs are first divided wherever there is a forward slash (/) and hyphen (-). Next, the domains, sub-domains, and extensions are separated by splitting wherever there is a dot (.). Lastly, the stop words are removed which is www and.com. The words in URLs are then assigned weights using TfidfVectorizer. TfidfVectorizer converts the text into feature vectors that is used as an input. TfidfVectorizer computes the word count, idf values and tf-idf score. If the words are unique, the score is high and common words have a low score.

## 3.4 Methodology

Different steps involved in methodology are given below:

1. Get the dataset.
2. Preprocess the data.
3. Divide the dataset into a training set and testing set.
4. Use different machine learning algorithms for training the model.
5. Get the evaluation matrices.
6. Get predictions on completely different data.
7. Make a classification table.

By default, the dataset is split into 75% training and 25% testing, and the further training set is converted into 75% training and 25% validation. The training set contains 85,098 samples, having 42,549 good and bad samples respectively, validation set consists of 28,366 samples divided into 14,183 good and bad samples, whereas the testing set has 37,822 samples of equally good and bad, respectively. In this experiment, initially, train the model using five supervised machine learning algorithms, viz. logistic regression algorithm, K-nearest neighbor algorithm, decision tree algorithm, support vector machine algorithm with 3 of its kernels, and random forest algorithm. Flowchart for designing individual algorithms is discussed in Fig. 3. Next, consider 10 models, each having 3 algorithms together to check the performance of those models. For ensemble models, the model is trained using three different algorithms of any combination. For calculating the accuracy and for generating the classification report for validation data, majority voting is calculated using the mode function. Each model makes predictions (votes) for every test event and the final outcome prediction is more than half the votes. If more than half of the votes were not cast in any of the predictions, it is not possible to predict a stable prediction for this event in a combined manner. Lastly, the model is given test data for prediction, and in a similar way, the accuracy and classification report for test data is obtained. The flowchart for the ensemble model is described in Fig. 4. After designing the algorithms, the classification based on four parameters namely, accuracy, precision, recall, and Fl-score was analyzed.

## 3.5 Performance Evaluation

### 3.5.1 Confusion Matrix

To get the idea about classifications and misclassifications, generate a confusion matrix for each algorithm. Confusion matrix is the m x m matrix, where m is number of variables. Table 1 shows the confusion matrix. Diagonal elements represent the true values and non-diagonal elements represent false values. There are four parameters inside the table, viz. False Positive (FP), True Negative (TN), True Positive (TP), and False Negative (FN).

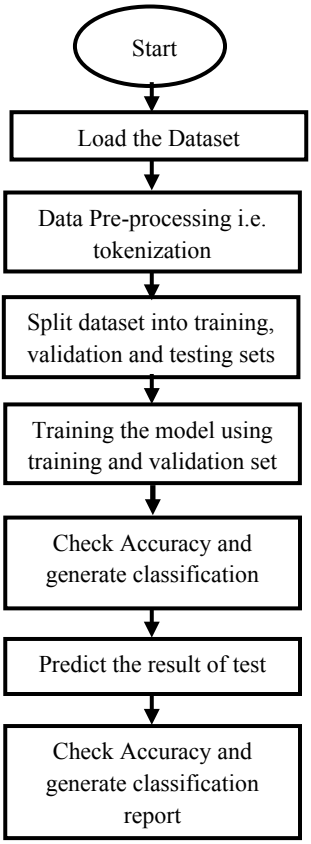



**Fig. 3** Flowchart of individual model

### 3.5.2 Accuracy

Accuracy is determined as the ratio of rightly predicted good and bad URLs to total prediction.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

### 3.5.3 Precision

Precision is determined as the ratio of literal good URLs out of total predicted good URLs.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

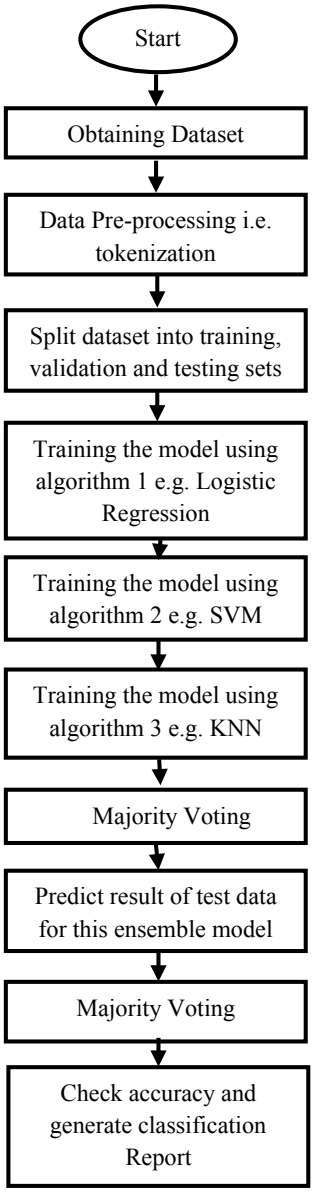



**Fig. 4** Flowchart of individual ensemble model

**Table 1** Confusion matrix

| Actual value | Predicted values | |
|---|---|---|
| | Negative | Positive |
| **Negative** | True negative | False positive |
| **Positive** | False negative | True positive |

### 3.5.4 Recall

The recall is defined as the ratio of actual good URLs over total actual good URLs.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

### 3.5.5 F1 Score

F1 scores are the harmonic mean of recall and precision. It is computed by the given formula:

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{4}$$

## 4 Results

All the individual algorithms were tried on different values of the random state. After implementing all the algorithms, the classification report of each algorithm is shown in Table 2, and the training–testing accuracy plots are depicted in Figs. 5 and 6.

From Figs. 5 and 6, it can be clearly seen that there is not much difference between training and testing accuracies which indicates that the models are not overfitted. Logistic regression algorithm has given the maximum accuracy 94.31%, and random forest algorithm has given the minimum accuracy of 87.34%. Precision, recall, and F1-score shown in Table 2 for each algorithm are above 80% indicating good classification between good and bad URLs.

**Table 2** Classification report of individual models

| Algorithm | Precision | | Recall | | F1-score | |
|---|---|---|---|---|---|---|
| | Bad | Good | Bad | Good | Bad | Good |
| Logistic regression | 0.93 | 0.95 | 0.95 | 0.94 | 0.94 | 0.94 |
| K-nearest neighbor | 0.97 | 0.82 | 0.84 | 0.96 | 0.90 | 0.89 |
| Decision tree | 0.91 | 0.87 | 0.88 | 0.90 | 0.89 | 0.89 |
| Random forest | 0.80 | 0.95 | 0.94 | 0.83 | 0.86 | 0.88 |
| SVM (linear) | 0.92 | 0.96 | 0.96 | 0.92 | 0.94 | 0.94 |
| SVM (rbf) | 0.90 | 0.97 | 0.96 | 0.91 | 0.93 | 0.94 |
| SVM (sigmoid) | 0.93 | 0.96 | 0.96 | 0.93 | 0.94 | 0.94 |

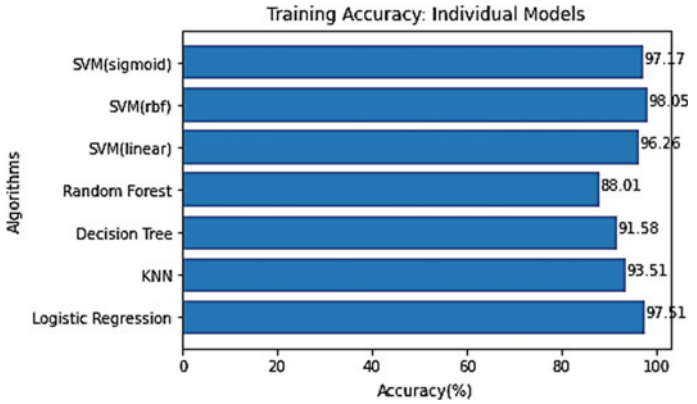



**Fig. 5** Training accuracy: individual models

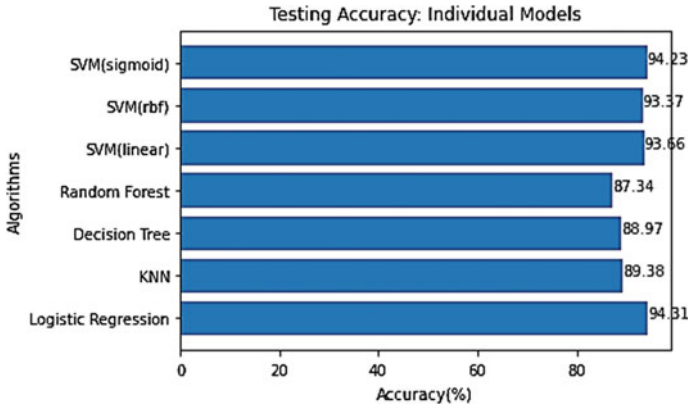



**Fig. 6** Testing accuracy: individual models

Next, combined all these three algorithms at a time. Since SVM (sigmoid kernel) gave a good result, this algorithm is selected instead of all the SVM kernels. The classification report of these algorithms is shown in Table 3, and the accuracies are depicted in Fig. 7.

From Fig. 7, it is clear that the eighth ensemble model (K-nearest neighbor, decision tree, and support vector machine) gives a maximum accuracy of 94.93%. The accuracy of all the ensemble models lies in the range 93–95. Also, the high (above 90%) precision, recall, and F1-score are shown in Table 3 which indicate that all of the ensemble models have performed well.

**Table 3** Classification report of ensemble models

| Algorithm | Precision | | Recall | | F1-score | |
|---|---|---|---|---|---|---|
| | Bad | Good | Bad | Good | Bad | Good |
| LR-KNN-DT | 0.95 | 0.95 | 0.95 | 0.94 | 0.95 | 0.95 |
| LR-KNN-RF | 0.96 | 0.93 | 0.93 | 0.96 | 0.94 | 0.94 |
| LR-KNN-SVM | 0.96 | 0.94 | 0.94 | 0.96 | 0.95 | 0.95 |
| LR-DT-RF | 0.95 | 0.92 | 0.92 | 0.95 | 0.94 | 0.94 |
| LR-DT-SVM | 0.97 | 0.94 | 0.94 | 0.97 | 0.95 | 0.96 |
| LR-RF-SVM | 0.96 | 0.93 | 0.93 | 0.96 | 0.94 | 0.94 |
| KNN-DT-RF | 0.94 | 0.93 | 0.93 | 0.94 | 0.94 | 0.94 |
| KNN-DT-SVM | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| KNN-RF-SVM | 0.96 | 0.93 | 0.93 | 0.96 | 0.94 | 0.95 |
| DF-RF-SVM | 0.95 | 0.92 | 0.92 | 0.96 | 0.94 | 0.94 |

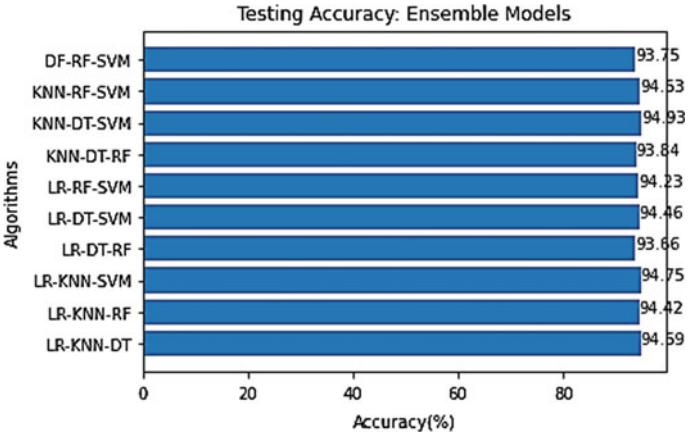



**Fig. 7** Testing accuracy: ensemble models

## 5 Conclusion

Malicious URL detection plays a significant role in cybersecurity applications. In this article, various ML algorithms are utilized to classify the URLs into malicious (bad) and legitimate (good). In the experiments, the performance accuracy of five different machine learning algorithms was compared. To get better results, the algorithms were implemented by changing different parameters like the kernel (in SVM algorithm), criterion and number_of_trees (in the decision tree and random forest algorithms), neighbors (in KNN algorithm), and random state for each individual algorithm. The same parameters were used to evaluate each ensemble model. By analyzing the experimental results, it is shown that our model has effectively classified the URLs.

Logistic regression and KNN-DT-SVM ensemble model have outperformed other algorithms. Overall, the accuracy of the individual model is above 85%, and the accuracy of each ensemble model is also above 92%. The major limitation of the proposed work is that the DT, RF and especially SVM takes hours for training. Future scope of this includes increasing the number of algorithms in the ensemble model and change parameters considering different features.

## References


1. Gaikwad S, Nale P, Bachate R (2016) Survey on big data analytics for digital world. In: IEEE international conference on advances in electronics, communication and computer technology. Pune, pp 180–186
2. Gabriel AD, Gavrilut DT, Alexandru BI, Stefan PA (2016) Detecting malicious URLs: a semi-supervised machine learning system approach. In: 18th ınternational symposium on symbolic and numeric algorithms for scientific computing. Timisoara, pp 233–239
3. Cybercrime statistics. https://niti.gov.in/sites/default/files/2019-07/CyberSecurityConclaveAtVigyanBhavanDelhi_1.pdf
4. Phishing statistics. https://www.thesslstore.com/blog/phishing-statistics-latest-phishing-stats-to-know/#:~:text=The%20latest%20estimate%20from%20ProofPoint's,email%20compromise%20(BEC)%20attacks
5. Coronavirus related spear phishing attacks. https://www.securitymagazine.com/articles/92157-coronavirus-related-spear-phishing-attacks-see-667-increase-in-march-2020
6. Alswailem A, Alabdullah B, Alrumayh N, Alsedrani A (2019) Detecting phishing websites using machine learning. In: 2nd ınternational conference on computer applications & ınformation security. Riyadh, Saudi Arabia, pp 1–6
7. Baraneetharan E (2020) Role of machine learning algorithms intrusion detection in WSNs: a survey. J Inf Technol Digital World 2:161–173
8. Ray S (2019) A quick review of machine learning algorithms. International conference on machine learning, big data, cloud and parallel computing. Faridabad, India, pp 35–39
9. Verma A, Mehta S (2017) A comparative study of ensemble learning methods for classification in bioinformatics. In: 7th international conference on cloud computing, data science & engineering—confluence. Noida, pp 155–158
10. Vanhoenshoven F, Nápoles G, Falcon R, Vanhoof K, Köppen M (2016) Detecting malicious URLs using machine learning techniques. In: IEEE symposium series on computational ıntelligence. Athens, pp 1–8
11. Tan G, Zhang T, Liu Q, Liu X, Zhu C, Dou F (2018) Adaptive malicious URL detection: learning in the presence of concept drifts. In: 17th IEEE ınternational conference on trust, security and privacy in computing and communications/12th IEEE ınternational conference on big data science and engineering (TrustCom/BigDataSE). New York, NY, pp 737–743
12. Feroz MN, Mengel S (2015) Phishing URL detection using URL ranking. In: IEEE ınternational congress on big data. New York, NY, pp 635–638
13. Manjeri AS, R K, V AMN, Nair PC (2019) A machine learning approach for detecting malicious websites using URL features. In: 3rd ınternational conference on electronics, communication, and aerospace technology. Coimbatore, India, pp 555–561
14. Kumar H, Gupta P, Mahapatra RP (2018) Protocol based ensemble classifier for malicious URL detection. 3rd ınternational conference on contemporary computing and ınformatics. Gurgaon, India, pp 331–336
15. Dataset. https://github.com/NetsecExplained/Machine-Learning-for-Security-Analysts

# Review on Energy-Efficient Routing Protocols in WSN


**G. Mohan Ram and E. Ilavarsan**

**Abstract** Recently, wireless sensor networks (WSNs) incorporate their prominent role in various applications like monitoring and tracking remote environments. WSN exhibits a distributed nature and dynamic topology which increases the challenge of designing an energy-efficient protocol for routing. Enhancement of energy efficiency in WSN is considered as the primary goal of the routing protocol. This review mainly intends to discuss the hierarchical-based energy-efficient routing protocols to maximize a lifetime of network and energy efficiency. The objectives, challenges, and issues of the WSN routing protocols are also discussed in this review. Finally, the performance analysis for each energy-efficient routing protocol is also summarized in this article. In this, the systematic literature survey from 2010 to 2020 for hierarchical-based energy-efficient routing protocol has been carried out. From these reviewed details, the researchers can obtain a valuable technical direction while emerging an energy-efficient routing protocol. The information available in this review is helpful for various researchers to acquire significant information about the current status of WSN's energy-efficient routing and the various potential concerns that need to be discussed. Finally, the future aspects and research gaps for the reviewed protocols are also discussed.

**Keywords** WSN · Routing protocol · Energy-efficient routing protocols · Hierarchical routing · Network lifetime

## 1 Introduction

As an evolving technology field for ad hoc networks, WSNs have been gaining extensive coverage. WSNs contain a variety of sensor nodes with wireless processing as well as communication proficiencies that are low cost, multi-functional, and low

G. M. Ram (✉) · E. Ilavarsan
Department of CSE, Pondicherry Engineering College, Puducherry, India

E. Ilavarsan
e-mail: eilavarsan@pec.edu



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_66

power. Such sensors connect within a small distance through a wireless intermediate and combine to achieve the common task, such as environmental detection, target management, and industrial process control [1]. The sensor network intends to use the collection of stationery, cheap, and tiny sensors for sensing the physical characteristics of the surrounding environment and to transfer such sensed information to its associated sink node [2]. However, WSN performs a variety of tasks among the most essential tasks like data sensing, data handling, and transferring the sensed data back to sink.

The routing protocol defines the information that allows the nodes to select the routes between the available nodes. In WSN, the routing process aims to send information to an acknowledged destination sensor node from a sensor node. The development of the efficient routing protocol achieves this objective by setting a path between the sensor and destination nodes. Few constraints in network resources such as link failures during communication, bandwidth, energy, meta-heuristic optimization problems, storage, and several other network constraints increase the complexity of designing an efficient routing protocol for WSN [3, 4]. The design of a routing protocol closely depends on the architecture model of an entire system.

During a network design process, energy conservation is a big cause of concern since sensors are fabricated with non-rechargeable batteries. When sending messages from source to destination, route selection is a crucial operation. During data transfer, vast amounts of energy are absorbed. The problem of unnecessary energy consumption is solved by using the extremely successful hierarchical routing architecture. The network is broken into many clusters in the clustering process. A middle sensor helps the source node to send information to an endpoint via routing tracks during the data transmission process [5, 6]. The principle of data aggregation helps in the successful utilization of available WSN resources.

The rest of the article is structured as follows. Section 2 describes the taxonomy of routing protocols in WSN. The energy-efficient routing protocols of WSN are described in Sect. 3. Section 4 explains the performance evaluation metrics of routing protocols. Section 5 describes the future enhancement of energy-efficient routing protocols. Finally, Sect. 6 represents the conclusion of routing protocols.

## 2 Classification of Routing Protocols in WSN

In modern decades, WSNs apply various routing protocols during data transmission to increase an entire network lifetime. Taxonomy of routing protocol is shown in Fig. 1. These protocols are categorized into two based on network structure and properties. Based on the node uniformity, the network structures are categorized. The primary feature of these protocol types is how the data is connected and exchanged by nodes depending on the interconnection mechanism.

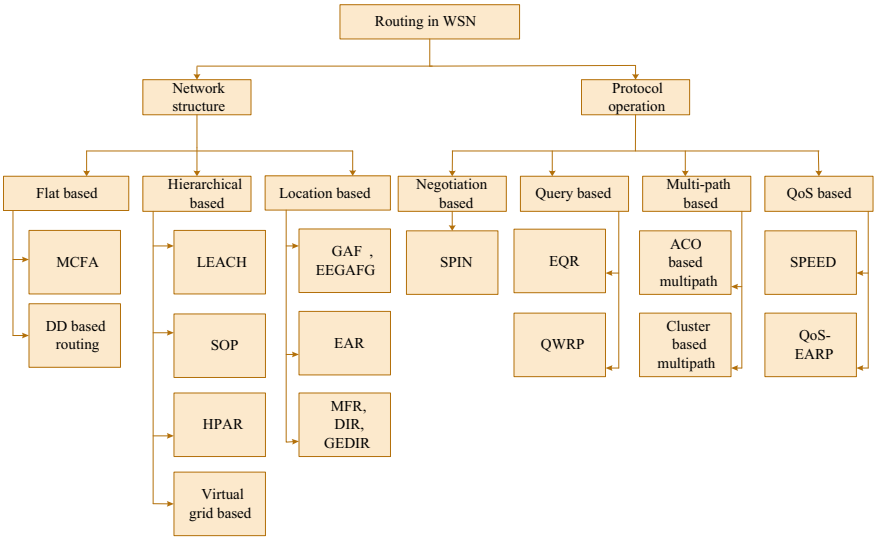



**Fig. 1** Taxonomy of routing protocols

## *2.1 Network Structure-Based Routing Protocols*

The hierarchical-based, location-based, and flat-based protocols come under the network structure-based routing protocols in which they are briefly explained in the given subsection.

### 2.1.1 Flat-Based Routing

This routing is necessary when massive numbers of sensor nodes are needed, where every single node performs a similar role. Therefore, it is not probable to allocate a special identifier (Id) for every single node if the number of nodes gets increased. Data-centric routing is led by this method in which the sink node forwards the information to a group of nodes for an answer. The nodes are treated impartially and have similar roles in this routing.

Minimum Cost Forwarding Algorithm (MCFA)

MCFA [7] takes advantage of the fact that the routing path is still remembered. For gradient-based routing, a sensor node does not need to have a specific ID. As an alternative, every node retains the smallest cost estimation from itself to the drain.

Any data sent by a sensor is sent to its neighbours. When the data is sent by a node, it checks if it is on a minimum expensive path between the source and the sink. It rebroadcasts the data to its neighbours if the path is least expensive. Before the sink is hit, this step occurs.

Directed Diffusion-Based

The hybrid energy balanced routing protocol (DCRP) and the diffusion clustering scheme (DCS) are proposed to avoid creating isolated CHs in the clustering process. Routing and clustering process are included in DCRP. DCS is projected in [8] to establish contact among neighbouring CHs without relay nodes during the clustering period. Clusters are created by diffusing hop by hop outward from the sink, while new CHs are picked from established clusters' member nodes. Besides, to minimize the number of clusters and transmission delays, the residual capacity, width, and number of neighbouring nodes outside the clusters have been taken into account. The hybrid energy balanced routing protocol (HEBR) is used in the routing process.

### 2.1.2 Hierarchical Cluster-Based Routing Protocol

In a hierarchical architecture, small-energy nodes may be utilized to measure the nearness of the destination node whereas higher energy nodes are utilized to route and relay signal facts. As a consequence, hierarchical cluster-based routing protocol strategies are reasonably effective for the sensor nodes within a cluster to consume fewer resources. Data aggregation and data fusion can also be achieved by these methods with a limited number of transferred messages to sink [9].

Low-Energy Adaptive Clustering Hierarchy (LEACH)

Energy is a main area of concern in WSNs due to lack of power supply. The highest capacity is required for data transfer. To maximize the lifespan of a sensor network, many experiments are performed. Among them, the solution focused on clustering is well established to accomplish energy efficiency. LEACH hierarchical routing protocol [10] has been proposed where clusters are regularly refreshed based on residual energy as well as space. Re-clustering distributes a capacity from numerous nodes, and in turn, the CH increases network lifespan. During its transmission slot only, the sensor nodes remain active. It lives in a sleepy state the majority of the time to conserve electricity.2.1.2.2 Self-Organizing Protocol (SOP).

SOP has been utilized to construct design to help diverse sensors [11]. These sensors may be stationary or mobile. Any sensor may measure the atmosphere and transfer the information to a specified group of nodes that serves as routers. Composed information is sent to the more efficient sink through the routers. A routing model has been suggested that involves the addressing of every single sensor. Sensing nodes may

be recognized by the address of the router node they are attached to. Local Markov loop (LML) algorithm was performed to promote fault tolerance and transmitting path. There are four stages in the algorithm for consolidating the router nodes and fabricating the routing tables, namely the exploration process, the organizational process, the management phase, and the self-reorganization phase.

Hierarchical Power-Aware Routing (HPAR)

Growth in wireless connectivity has made the development of low-cost WSNs feasible in recent years. An efficient topology management strategy is clustering sensor nodes. A new routing algorithm that can maximize the lifespan of the network was suggested in [12]. Each node will approximate its residual energy and then suggest a new form of clustering to maximize the existence of the network. This inference is identical to several other suggested WSN routing algorithms. Based on particular threshold values, the pre-defined numbers of nodes with the highest residual energy are initially chosen in the current algorithm as CHs and the representatives of every cluster are calculated based on the distance between the node and CH as well as between the CH and sink.

Virtual Grid Architecture

Since it has a basic and hustle-free design, the wireless network holds a special place in networking, less costly to save time and resources in numerous ways. But as they are human un-attended, therefore this resource-scare network needs special management. WSN still required modern routing and new technical developments in which energy in the network is used for data transmission. The two main problems in the WSNs are the utilization of wireless transmission energy and coverage. [13] proposes a routing algorithm that retains network coverage as well as reduces energy consumption, contributing to increased network life.

### 2.1.3 Location-Based Routing

Sensor nodes are uniformly distributed in the interesting regions in this form of network architecture and are often identified by the geographical location. They are mainly found by way of GPS. The separation between nodes is determined by the signal intensity generated from those nodes, and the coordinates are measured by the information exchange among adjacent nodes. Location-based routing protocols are utilized to identify nodes about their current location. The position is derived from GPS or by collaboration between nodes.

Geographical Adaptive Fidelity (GAF)

GAF is a routing protocol based on positions and energy focused. It was previously recommended for ad hoc networks, but now this protocol can also be used for WSNs. It is based on the principle that the perspective of routing is equal to all adjacent nodes. The entire network in GAF is broken into a virtual grid. Grid size is based on the idea in which every node can connect with the other node's adjacent grids. Each node utilizes GPS-based position information to link itself within the grid. This protocol depends on location, but is often utilized as a hierarchical protocol in which clusters are generated based on the location data [14].

Geographical and Energy-Aware Routing (GEAR)

For WSNs, a centralized clustering of regional energy-conscious routing (GEAR-CC) was introduced in [15]. It has the superiority of both hierarchical routing and geographic routing of spatial energy. It takes full advantage of the adequate capacity of the base station as a centralized algorithm to execute any transmission in the WSN. The base station can easily devise the optimum transmission schemes for all sensor nodes in GEAR-CC, based on global topology and energy details. The optimization is accomplished by generating trade-offs between the expense of electricity and the remaining capacity of the node.

MFR, DIR, GEDIR

Both the neighbours of the sender node are involved in routing decisions with directional routing protocol (DIR), most forwarded within range (MFR), and geographic distance routing (GEDIR) techniques [16]. Owing to the presence of the nodes lying in a backward direction, this adds to needless overhead and excessive energy consumption. This extra energy consumption minimizes the reduction of network lifespan. A new approach is recommended in [17], and a forward search area is added where a large number of sensors engage in routing decision-making.

## *2.2 Protocol Operation-Based Routing Protocols*

It is a different routing functionality. It is categorized into 4 types; they are QoS-based routing, query-based routing, multipath routing protocols, and negotiation-based routing protocols.

### 2.2.1 Multipath Routing Protocols

These protocols utilize multiple paths for network throughput improvement. It comprises two different ways for performing the transmission between the base station and source sensors.

(i) Disjoint paths
(ii) Braided paths.

For WSN, Yang [18] developed an ant colony optimization (ACO)-based multi-path routing protocol. CH is designated from the cluster of sensors based on residual energy. After that, a multipath between sink and CH is discovered by the ACO procedure. Based on the energy consumption parameter, CH dynamically chooses the route to convey the information with a probability. Cluster-based multi-path routing protocol was developed by Sharma and Jana [19] for WSN. Minimization of energy consumption as well as maximization of reliability is obtained by this clustering and multipath routing method. The major objective of this article is to minimize a load of the sensor by applying more accountability to the sink node.

### 2.2.2 Query-Based Protocol

Query-based protocol utilizes target node broadcast query for information. These queries are explained in terms of directed diffusion and rumour routing protocol.

Ahvar et al. [20] developed a new energy-aware query-based routing (EQR) for WSN. This EQR protocol provides a good transaction among the energy-saving objects as well as traditional energy balancing methods. The total energy consumption is decreased with the help of learning automata together with the zonal broadcasting method. For WSN with mobile sink, Jain et al. [21] developed query-based routing protocol. For query-driven scenarios, this paper developed a query-driven virtual wheel-based routing protocol (QWRP) method which creates a virtual structure with the aim of confining the sink node location. To improve the data delivery performance, a novel packet forwarding mechanism is developed under QWRP and an angle-based forwarding algorithm.

### 2.2.3 QoS-Based Routing Protocol

This protocol maintains a perfect balance between energy consumption and data quality. Few of its types are SPEED and energy-aware QoS routing.

Fonoage et al. [22] developed a QoS-based routing protocol. To forward packets in the network, this article developed a geographic routing mechanism combined with QoS support. The data is routed to other nodes based on the type of data packet. Here, multiple transmission queues are used to route packets with different significances. The node with high link quality, low load, and high residual energy is designated as a next-hop node which is nearer to sink node.

In WSN, Masruroh and Khadijah [23] developed a QoS-based and emergency-aware routing protocol (QoS-EARP). This article is developed based on the initial network design. Sensor network's emergency awareness, QoS, and energy efficiency are considered for algorithm design.

#### 2.2.4 Negotiation-Based Routing Protocol

High-level descriptors are used by negotiation-based routing protocol for removing the redundant data transmissions via negotiation.

Sensor Protocol for Information Negotiation (SPIN)

Data transmission is considered as one of the main tasks in WSN. Several routing protocols were suggested to conserve energy during data transmission. In this sense, data-centric-based routing methods are ideal for conducting in-network data aggregation to generate energy-saving data propagation. In [24], an updated variant of the SPIN protocol called M-SPIN was introduced and equates its efficiency with the standard SPIN protocol using broadcast communication, which is a well-known benchmark protocol. In the TOSSIM setting, the M-SPIN protocol is tested using simulation. M-SPIN has major efficiency benefits relative to conventional SPIN routing.

## 3 Energy-Efficient Routing Protocols in WSN

In the case of wireless networks, battery-operated devices try to achieve heuristic energy efficiency. Conversely, this is not the optimal strategy for multi-hop routing, which would be typical for sensor and ad hoc networks. In this review, the following types of energy-efficient routing protocols were analysed and established out in Fig. 2.

### *3.1 Opportunistic Routing*

Opportunistic routing is a new revolution for WSN routing that decides the sensor closest to the target sensor for data transmission. It utilizes the nature of the transmission of wireless communication. It enhances the effectiveness, presentation, and dependability of a network. So, this section presents a brief explanation of the opportunistic routing protocol in WSN (Table 1).

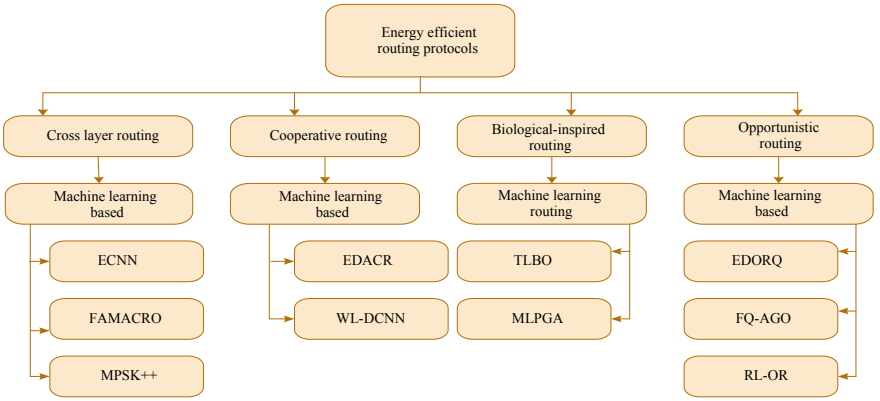



**Fig. 2** Classification of energy-efficient routing protocols

**Table 1** Survey on opportunistic routing

| Author name and year | Protocol name | Objectives | Advantages | Disadvantages | Metrics | Future scope |
|---|---|---|---|---|---|---|
| Zhu and Don Towsley [25] | $E^2R$ | To deliver control messages and data packets in a multi-hop wireless network | Node failure percentage is less | Causes poor end-to-end performance due to unavailability of preselected routing paths | Packet delivery ratio, control overhead, packet delivery delay | – |
| Hung et al. [26] | EFFORT | To maximize the amount of data gathered in WSNs | Improves transmission reliability | Cannot adopt per-hop communications due to the end-to-end data transmission | Network lifetime, end-to-end delay, energy cost, error percentage | – |
| Devi et al. [27] | EESOR: | To reduce delay in transmission and to prolong the network lifetime | Increase reliability and improve the sensing range of the sensor | More number of hops increases the delay in data transmission | Network lifetime and end-to-end delay | Analyse this routing protocol for throughput and turnaround time parameters |
| Chithaluru et al. [28] | AREOR | An energy-efficient optimal forwarder node selection | Using the sensor's for energy while assuring the QoS parameter | Due to the sub-optimal selection of routes, there is a loss in throughput | Message success rate, energy consumption, end-to-end delay, and packet delivery ratio | Fine-tuning of the adaptive method to decide the node rank quickly |

### 3.2 Cross-layer Routing

In WSN, there is a need for cooperation among the various layers in the protocol stack for routing problems to conserve energy and to improve application performance. Cross-layer architecture accomplishes this kind of routing. This architecture enables different layers to collaborate and share network position data and guarantees the best route which is chosen for better energy enhancement purpose (Table 2).

### 3.3 Cooperative Routing

Cooperative routing is a kind of cross-layer routing that uses cooperative communication technology in the physical layer and routing selection in the network layer. This routing is mainly performed to decrease the energy consumption in the overall network (Table 3).

### 3.4 Biological Inspired Optimal Routing

The self-organization and self-association characteristics are more desirable by swarm intelligence (SI)-based protocols, and besides it contributes to organizing both the negative and the positive feedbacks. The routing behaviour is mainly carried out by certain different SI-assisted meta-heuristic algorithms such as bee colony optimization (BCO), ACO, and particle swarm optimization (PSO). In ACO, indirect coordination called stigmergy is followed by ants, and this approach helps to solve the combination issues. Stigmergic communication is completely fulfilled due to the compound substances gained from the ants called pheromone. Besides, such pheromones from the ants also act as a real feedback tool, so more ants are enrolled to attain their ultimate destination by following the pheromones which are sequenced on the shortest path (Table 4).

### 3.5 Machine Learning (ML)-Based Routing Techniques Under Opportunistic, Cross-layer, Swarm Intelligence, and Cooperative Protocols

This section briefly explains the machine learning algorithm which comes under the opportunistic, cross-layer, swarm intelligence, and cooperative routing protocols. In next-generation network, machine learning-based routing algorithms show great

**Table 2** Survey on cross-layer routing

| Author name and year | Protocol name | Objectives | Advantages | Disadvantages | Future scope | Metrics |
|---|---|---|---|---|---|---|
| Singh and Verma [29] | Adaptive threshold-based routing protocol | To perform routing in heterogeneous networks | Increase the lifetime of the whole network | Data overhead is high | – | Network lifetime and number of alive nodes |
| Ward and Younis [30] | Distributed beamforming-based routing | Balances the relay recruitment energy and the number of recruited relays | Improve the energy efficiency of the whole network | Data traffic is not limited | Enhancing the non-distributed approach in the WSN field | Average communication energy |
| Han et al. [31] | Geographic node disjoint multipath routing protocol | To optimize the whole network | Optimize all the parameters to achieve routing | Limits the useable links for exploring possible routing paths | Selecting path with minimum inter-interference | Residual energy, sum rate, and available path length |
| Abazeed et al. [32] | CLMR | To ensure QoS and minimize energy consumption | Reliability, security, energy efficiency, fault tolerance | Due to the memory constraint, the neighbour table is limited with few numbers of nodes | Integration of CLMR with Internet | Packet delivery ratio, PSNR, end-to-end delay |

**Table 3** Survey on cooperative routing

| Author name and year | Protocol name | Objectives | Advantages | Disadvantages | Future scope | Metrics |
|---|---|---|---|---|---|---|
| Hung et al. [33] | EERH | To reduce the energy consumption of the whole network | Reduce the energy consumption of the whole network | Limited packet size | Security of EERH to be achieved | Average residual energy |
| Rani et al. [34] | CBCCP | To minimize the transmission distance | Improve network lifetime | Communication gets delayed due to limited power | Security must be considered | Delay and number of dead nodes |
| Manfredi [35] | Cooperative routing protocol | To guarantee a good performance trade-off between reliabilities | Energy efficiency enhancement | The limited capacity of the wireless medium network | To adapt a cooperative strategy for particular applications | Energy consumption and time delay |
| Habibi et al. [36] | Mixed-integer optimization framework | To obtain minimum energy transmission policy | Reduces energy consumption | Less amount of computational capability | Security must be considered in future studies | Bit error rate and normalized transmission power |
| Razzaque et al. [37] | DACR | To reduce delay and energy | Provides reliability of the network | Lack of understanding about the dynamics of several estimation tuning parameters | The analytical model must be performed in future | Energy consumption, packet delivery ratio, end-to-end delay |

**Table 4** Survey on biological inspired optimal routing

| Author name and year | Protocol name | Objectives | Advantages | Disadvantages | Future scope | Metrics |
|---|---|---|---|---|---|---|
| Helmy et al. [38] | AFSA | To lower down network energy consumption | Achieves better network lifetime | Energy consumption is somewhat high | Limiting the energy of the nodes that have little or no energy is another area of future research | FND, energy consumption per round, network lifetime, data received by BS |
| Liu et al. [39] | QoS-PSO | To enhance the QoS level | Improvement in the QoS measure | Due to limited resources, maintenance of data structure was affected | Improvement to effectively handle high mobility | Packet loss, average residual energy, QoS, mean delay |
| Kuila and Jana [40] | PSO | To extend lifetime of WSNs | Energy consumption is balanced, and network lifetime is improved | Overhead of data routing in the CH formation phase | Enhancement to deal with dynamic network scenarios | Network lifetime, energy consumption, delivery of total data packets |
| Rao et al. [41] | PSO-ECHS | To achieve improved network lifetime by conserving the energy of SNs | Better performance in terms of success rate, network lifespan, and total energy consumption | Inappropriate cluster formation | Fault tolerance in heterogeneous networks | Total energy consumption, network lifetime |
| Shankar et al. [42] | HSA-PSO | To attain balanced energy consumption among SNs | Achieves better search, convergence, energy-efficient operation | Inappropriate cluster formation | Scalability enhancement for very large-scale networks | FND, LND, residual energy, mean throughput, standard deviation |

prospective conversely, and researches on artificial intelligent routing are still on a very beginning stage. So, this section briefly explains the machine learning concepts in routing protocols (Table 5).

## 4 Performance Evaluation Metrics of Routing Protocols

For performance analysis, this section shows the various performance metrics of routing protocols. Some of the parameters were selected to make an adequate approach to the technical specifications of the routing protocols.

1. Throughput: The amount of bit transmission per second is defined as throughput, and it is measured by Kbps. Such type of calculation provides a fast response about system productivity boosted by every group.
2. End-to-end delay: It is a period taken by the standard data transmission system from the source node to the target node.
3. Normalized overhead routing: It is described as the number of packets transmitted per data packet.
4. Packet delivery ratio: It is defined as the ratio of several packets received by the destination node to the number of packets sent. It is measured in terms of percentage.

The given Tables 6, 7, 8, 9, and 10 represent the various performance parameter values of different types of energy-efficient routing protocols.

## 5 Future Enhancement

In the proposed work, future studies should focus on the design and implementation of unique routing protocols for particular applications. To develop routing protocols for applications needing QoS services such as real-time applications, video, imaging, and surveillance monitoring, careful attention is required. In this field, some of the research processes are performed so far to explore a vast. The survey revealed that protocol expectations on SNs and sinks are typically stationary. It is typically not a real-environment scenario, though. Applications such as goal detection and analysis of surveillance serve as a mobility prerequisite. Further study is needed to explore the effectiveness of routing protocols to accommodate volatility of sink/source nodes, the overhead due to complex topology, QoS constraint in an area of energy restriction. It is important to facilitate the algorithm experimentation on actual testbeds so there could be some elements of a protocol that can be discovered there and not discovered on the simulator. It is expected to inspire researchers to consider different characteristics of the protocol when designing and improving energy-efficient routing protocols that are energy efficiency, application area, statistical modelling, QoS calculation, and implementation of simulation/real testbed.

**Table 5** Survey of machine learning-based routing methods in energy-efficient routing protocols

| Author name and year | Type of routing | Protocol name | Advantages | Disadvantages | Future scope | Metrics |
|---|---|---|---|---|---|---|
| Sumalatha and Nandalal [43] | Cross-layer | ECNN | Reduces faults in routing | Packets are misrouted | To find the various kinds of attacks with different parameters | Throughput, energy consumption |
| Gajjar et al. [44] | Cross-layer | FAMACRO | Reduces collision among the clusters | Limited hostile environment | – | Energy consumption and residual energy |
| Mydhili et al. [45] | Cross-layer | MSPK + + | Eliminates the corrupted data | Energy consumption is high | Optimized the ML through the heuristic approach | Accuracy, computational time |
| Yongjie Lu et al. [46] | Opportunistic routing | EDORQ | Improves network performance | Overhead of data | Minimize the average packet delay | Network overhead, packet delay, packet delivery ratio, energy consumption |
| Alshehri et al. [47] | Opportunistic routing | FQ-AGO | Selects the best candidate set towards the destination | Large overhead | – | Packet delivery ratio, throughput, delay |
| Tang et al. [48] | Opportunistic routing | RL-OR | Improves transmission reliability | Limited wireless resources | – | End-to-end delay, throughput |
| Wang et al. [49] | Cooperative routing | EDACR | Adjusts the weight for three metrics | Non-supportive behaviour for real-time traffic | – | Average QoS, percentage of surviving nodes |
| Huang et al. [50] | Cooperative routing | WL-DCNN | Determines the reliability of target links | Not sensitive to data with spatial displacement | Apply this technique to improve the resilience of the network | Average routing length, number of dead nodes |

**Table 6** Comparison values of opportunistic routing protocol

| Reference no. | Method | Parameters | Values |
|---|---|---|---|
| Chithaluru et al. [28] | AREOR | Packet delivery ratio | 97.78% |
| | | End-to-end delay | 36.6 s |
| | | Message success rate | 230 bits/s |
| | | Consumed energy (CE) | 72 J |
| Zhu and Towsley [25] | $E^2R$ | Packet delivery ratio | 100% |
| | | Control overhead | 0.4 |
| | | Packet delivery delay | 1.4 s |
| Hung et al. [26] | EFFORT | Network lifetime | 2 MB |
| | | Energy cost | 0.012 mJ/bit |
| | | End-to-end delay | 20.5 s |
| | | Error percentage | 0.12% |
| Devi et al. [27] | EESOR | Average end-to-end delay | 500 |
| | | Maximum end-to-end delay | 800 |
| | | Network lifetime | 127.5 |

**Table 7** Comparison values of cross-layer routing protocol

| Reference no. | Method | Parameters | Values |
|---|---|---|---|
| Singh and Verma [29] | Adaptive threshold-based routing protocol | Energy consumption | 66% |
| | | Number of alive nodes | 10@80 rounds |
| Ward and Younis [30] | DiCLR | Average communication energy | 3 μJ |
| | | Belief metric | 0.01 |
| Han et al. [31] | Geographic node disjoint multipath routing protocol | Average residual energy | 490 |
| | | Average length of paths | 24 |
| Abazeed et al. [32] | CLMR | Delivery ratio | 0.94% |
| | | Energy per packet | 0.065 |

Future studies will help researchers to attain reasonable decisions and to set targets based on better options for investigations.

## 6 Conclusion

WSN becomes more common in recent years; therefore, it is extensively used in various applications like volcano or earthquake prediction, health care, structural health monitoring, target tracking, remote sensing, intruder detection, surveillance, and military. Due to the restricted battery capacity of source nodes, energy usage

**Table 8** Comparison values of cooperative routing protocol

| Reference no. | Method | Parameters | Values |
|---|---|---|---|
| Hung et al. [33] | EERH | Average residual energy | 95 mJ |
| | | Average notified event | $13.5 \times 10^4$ |
| Rani et al. [34] | CBCCP | Average number of alive nodes | 50@5000 rounds |
| | | Number of dead nodes | 1100@5000 rounds |
| Manfredi [35] | Cooperative routing protocol | Delay | 3.6 s |
| | | Reliability | 68% |
| Habibi et al. [36] | Mixed-integer optimization framework | BER | 0.038 |
| | | Normalized transmission power | 37.5 |
| Razzaque et al. [37] | DACR | Average end-to-end delay | 0.3 s |
| | | Network lifetime | 400 |

**Table 9** Comparison for optimization-based energy-efficient routing protocol

| Reference no. | Method | Parameters | Values |
|---|---|---|---|
| Helmy et al. [38] | AFSA | Number of alive nodes | 45@100 rounds |
| | | Residual energy mean value | 42 J |
| Liu et al. [39] | QoS-PSO | Mean delay | 3.2 s@100 nodes |
| | | Packet loss ratio | 0.38 @100 nodes |
| | | QoS | 0.58@100 nodes |
| | | Average residual energy | 0.9 J for 100 nodes |
| Kuila and Jana [40] | PSO | Network lifetime | 600 nodes @ 650 rounds |
| | | Inactive rate of sensor node | 135@2500 rounds |
| | | Energy consumption | 1650 J@3000 rounds |
| | | Packet received by BS | $7.56 \times 10^4$ for 90BS |
| Rao et al. [41] | PSO-ECHS | Total energy consumption | 450 J@ 5000 rounds |
| | | Packet received by BS | 540 for 15 CHs |
| Shankar et al. [42] | HSA-PSO | Mean residual energy | 98 J |
| | | Mean throughput | 17 Mbps for 200-100 sink positions |

is a big issue for these devices. Hardware constraints are required for designing an energy-efficient hierarchical routing protocol. This article highly attempts to review the classical- and hierarchical-based routing protocols that intend to progress the efficiency of network energy. In-depth empirical comparison for various hierarchical-based routing protocols is provided in this article. Further, the performance can be measured by the comparing the various parameters such as QoS, energy consumption,

**Table 10** Comparison of machine learning-based energy-efficient routing protocol

| Reference no. | Method | Parameters | Values |
|---|---|---|---|
| Sumalatha and Nandalal [43] | ECNN | Throughput | 0.1034 bps@5.4 s |
| | | Energy consumption | 9.90 J@10 s |
| Gajjar et al. [44] | FAMACRO | Average energy consumption | 20.240 J@250 rounds |
| | | Number of dead nodes | 1000@250 rounds |
| | | Sum of residual energy | 270 J@250 rounds |
| Mydhili et al. [45] | MSPK + + | Residual energy | 85.56 J@100 nodes |
| | | Computational time | 0.3 s@100 nodes |
| Yongjie Lu et al. [46] | EDORQ | Total energy consumption | 7.8 kJ@800 nodes |
| | | Packet delivery ratio | 0.98@800 nodes |
| | | Average packet delay | 1.8 s@800 nodes |
| Alshehri et al. [47] | FQ-AGO | Throughput | 1600kbps@30 m/s |
| | | Delay | 0.4 s@30 m/s |
| | | Packet delivery ratio | 0.75%@30 m/s |
| Tang et al. [48] | RL-OR | End-to-end delay | 0.1 s@5Mbps |
| | | Throughput | 5 mbps@10 s delay |
| Wang et al. [49] | EDACR | QoS | 0.38@3000 rounds |
| Huang et al. [50] | WL-DCNN | Accuracy | 0.986 for yeast, 0.971 for power |

fault tolerance, question regarding multipath, scalability, position knowledge, and data aggregation were also analysed. The classical routing techniques that stress the load balancing evolutionary techniques and optimal clustering are mainly reviewed in this work to find out a strong correlation between the network lifetime and energy efficiency. Through proposing architecture, algorithmic level improvements by biologically driven meta-heuristic methods that provide better solutions for optimization issues, hierarchical routing in terms of swarm intelligence enhances a reasonable contribution. This study provides a better solution for the problems related to energy consumption for hierarchical-based routing protocols; further, it contributes to the better network lifespan. Finally, the open challenges and issues that are encountered while designing an energy-efficient routing protocol are also outlined in this work.

## References


1. BenSaleh MS, Saida R, Kacem YH, Abid M (2020) Wireless sensor network design methodologies: a survey. J Sens
2. Othman MF, Shazali K (2012) Wireless sensor network applications: a study in environment monitoring system. Procedia Eng 41:1204–1210
3. Shabbir, N, Hassan SR (2017) Routing protocols for wireless sensor networks (WSNs). Wirel Sens Netw Insights Innov
4. Ketshabetswe LK, Zungeru AM, Mangwala M, Chuma JM, Sigweni B (2019) Communication protocols for wireless sensor networks: a survey and comparison. Heliyon 5(5):e01591
5. Engmann F, Katsriku FA, Abdulai JD, Adu-Manu KS, Banaseka FK (2018) Prolonging the lifetime of wireless sensor networks: a review of current techniques. Wirel Commun Mobile Comput
6. Sabor, N., Sasaki, S., Abo-Zahhad, M, Ahmed SM (2017) A comprehensive survey on hierarchical-based routing protocols for mobile wireless sensor networks: review, taxonomy, and future directions. Wirel Commun Mobile Comput
7. Cecílio J, Costa J, Furtado P (2010) Survey on data routing in wireless sensor networks. In Wireless sensor network technologies for the information explosion era. Springer, Berlin, Heidelberg, pp 3–46
8. Yinghong L, Yuanming W, Jianyu C (2019) The diffusion clustering scheme and hybrid energy balanced routing protocol (DCRP) in multi-hop wireless sensor networks. Adhoc Sensor Wirel Netw 43
9. Singh SK, Singh MP, Singh DK (2010) A survey of energy-efficient hierarchical cluster-based routing in wireless sensor networks. Int J Adv Netw Appl (IJANA) 2(2):570–580
10. Singh J, Singh BP, Shaw S (2014, September) A new LEACH-based routing protocol for energy optimization in wireless sensor network. In: 2014 international conference on computer and communication technology (ICCCT). IEEE, pp 181–186
11. Chaubey NK, Patel DH (2016) Routing protocols in wireless sensor network: a critical survey and comparison. Int J IT Eng 4(2):8–18
12. Golsorkhtabar M, Hosinzadeh M, Heydari MJ, Rasouli S (2010) New power aware energy adaptive protocol with hierarchical clustering for WSN. Int J Comput Netw Sec 2(4):38–40
13. Jain KL, Mohapatra S (2019) Grid base energy efficient coverage aware routing protocol for wireless sensor network. In: Proceedings of the 2nd international conference on software engineering and information management, pp 49–53
14. Grover J, Sharma M (2014) Location based protocols in wireless sensor network—a review. In: Fifth international conference on computing, communications and networking technologies (ICCCNT). IEEE, pp 1–5
15. Tang B, Wang D, Zhang H (2013, October) A centralized clustering geographic energy aware routing for wireless sensor networks. In: 2013 IEEE international conference on systems, man, and cybernetics. IEEE, pp 1–6
16. Kumar V, Kumar S (2016) Energy balanced position-based routing for lifetime maximization of wireless sensor networks. Ad Hoc Netw 52:117–129
17. Priya IL, Lalitha S, Paul PV (2018) Energy efficient routing models in wireless sensor networks-a recent trend survey. Int J Pure Appl Math 118(16):443–458
18. Yang J, Xu M, Zhao W, Xu B (2010) A multipath routing protocol based on clustering and ant colony optimization for wireless sensor networks. Sensors 10(5):4521–4540
19. Sharma S, Jena SK (2015, Apr 22) Cluster based multipath routing protocol for wireless sensor networks. ACM SIGCOMM Comput Commun Rev 45(2):14–20
20. Ahvar E, Serral-Gracià R, Marín-Tordera E, Masip-Bruin X, Yannuzzi M (2012, June) EQR: a new energy-aware query-based routing protocol for wireless sensor networks. In: International conference on wired/wireless internet communications. Springer, Berlin, Heidelberg, pp 102–113
21. Jain S, Pattanaik KK, Shukla A (2019) QWRP: query-driven virtual wheel based routing protocol for wireless sensor networks with mobile sink. J Netw Comput Appl 147:102430

22. Fonoage M, Cardei M, Ambrose A (2010, December). A QoS based routing protocol for wireless sensor networks. In: International performance computing and communications conference. IEEE, pp 122–129
23. Masruroh SU, Sabran KU (2014) August. Emergency-aware and QoS based routing protocol in wireless sensor network. In: 2014 international conference on intelligent autonomous agents, networks and systems. IEEE, pp 47–51
24. Rehena Z, Roy S, Mukherjee N (2011, January). A modified SPIN for wireless sensor networks. In 2011 third international conference on communication systems and networks (COMSNETS 2011). IEEE, pp 1–4
25. Zhu T, Towsley D (2011, April) E 2 R: energy efficient routing for multi-hop green wireless networks. In: 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE, pp 265–270
26. Hung MCC, Lin KCJ, Chou CF, Hsu CC (2013) EFFORT: energy-efficient opportunistic routing technology in wireless sensor networks. Wirel Commun Mobile Comput 13(8):760–773
27. Devi CY, Shivaraj B, Manjula SH, Venugopal KR, Patnaik LM (2014, March) EESOR: energy efficient selective opportunistic routing in wireless sensor networks. In: International conference on security in computer networks and distributed systems. Springer, Berlin, Heidelberg, pp 16–31
28. Chithaluru P, Tiwari R, Kumar K (2019) AREOR–adaptive ranking based energy efficient opportunistic routing scheme in wireless sensor network. Comput Netw 162:106863
29. Singh R, Verma AK (2017) Energy efficient cross-layer based adaptive threshold routing protocol for WSN. AEU-Int J Electron Commun 72:166–173
30. Ward JR, Younis M (2016, December) An energy-efficient cross-layer routing approach for wireless sensor networks using distributed beamforming. In: 2016 IEEE global communications conference (GLOBECOM). IEEE, pp 1–6
31. Han G, Dong Y, Guo H, Shu L, Wu D (2015) Cross-layer optimized routing in wireless sensor networks with duty cycle and energy harvesting. Wirel Commun Mobile Comput 15(16):1957–1981
32. Abazeed M (2019) Cross-layer multipath routing scheme for wireless multimedia sensor network. Wirel Netw 25(8):4887–4901
33. Hung LL, Leu FY, Tsai KL, Ko CY (2020) Energy-efficient cooperative routing scheme for heterogeneous wireless sensor networks. IEEE Access 8:56321–56332
34. Rani S, Malhotra J, Talwar R (2015) Energy efficient chain based cooperative routing protocol for WSN. Appl Soft Comput 35:386–397
35. Manfredi S (2012) Reliable and energy-efficient cooperative routing algorithm for wireless monitoring systems. IET Wirel Sens Syst 2(2):128–135
36. Habibi J, Ghrayeb A, Aghdam AG (2013) Energy-efficient cooperative routing in wireless sensor networks: a mixed-integer optimization framework and explicit solution. IEEE Trans Commun 61(8):3424–3437
37. Razzaque MA, Ahmed MHU, Hong CS, Lee S (2014) QoS-aware distributed adaptive cooperative routing in wireless sensor networks. Ad Hoc Netw 19:28–42
38. Helmy AO, Ahmed S, Hassenian AE (2015) Artificial fish swarm algorithm for energy-efficient routing technique. Intelligent systems 2014. Springer, Berlin, pp 509–519
39. Liu M, Xu S, Sun S (2012) An agent-assisted QoS-based routing algorithm for wireless sensor networks. J Netw Comput Appl 35(1):29–36
40. Kuila P, Jana PK (2014) Energy efficient clustering and routing algorithms for wireless sensor networks: particle swarm optimization approach. Eng Appl Artif Intell 33:127–140
41. Rao PS, Jana PK, Banka H (2016) A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks. Wirel Netw 1–16
42. Shankar T, Shanmugavel S, Rajesh A (2016) Hybrid HSA and PSO algorithm for energy efficient cluster head selection in wireless sensor networks. Swarm Evol Comput 30:1–10
43. Sumalatha MS, Nandalal V (2020, Mar 2) An intelligent cross-layer security based fuzzy trust calculation mechanism (CLS-FTCM) for securing wireless sensor network (WSN). J Ambient Intell Humanized Comput 1–5

44. Gajjar S, Sarkar M, Dasgupta K (2015, Jan 1) FAMACRO: fuzzy and ant colony optimization based MAC/routing cross-layer protocol for wireless sensor networks. Procedia Comput Sci 46:1014–1021
45. Mydhili SK, Periyanayagi S, Baskar S, Shakeel PM, Hariharan PR (2019, September 11) Machine learning based multi scale parallel K-means++ clustering for cloud assisted internet of things. Peer-to-Peer Netw Appl 1–3
46. Lu Y, He R, Chen X, Lin B, Yu C (2020, January) Energy-efficient depth-based opportunistic routing with Q-learning for underwater wireless sensor networks. Sensors 20(4):1025
47. Alshehri A, Badawy AH, Huang H (2020, April) FQ-AGO: fuzzy logic Q-learning based asymmetric link aware and geographic opportunistic routing scheme for MANETs. Electronics 9(4):576
48. Tang K, Li C, Xiong H, Zou J, Frossard P (2017, October 16) Reinforcement learning-based opportunistic routing for live video streaming over multi-hop wireless networks. In: 2017 IEEE 19th international workshop on multimedia signal processing (MMSP). IEEE, pp 1–6
49. Wang D, Liu J, Yao D, Member IE (2020, May 24) An energy-efficient distributed adaptive cooperative routing based on reinforcement learning in wireless multimedia sensor networks. Comput Netw 107313
50. Huang R, Ma L, Zhai G, He J, Chu X, Yan H (2020, Mar 31) Resilient routing mechanism for wireless sensor networks with deep learning link reliability prediction. IEEE Access 8:64857–64872

# Intelligent Machine Learning Approach for CIDS—Cloud Intrusion Detection System


**T. Sowmya and G. Muneeswari**

**Abstract** In this new era of information technology world, security in cloud computing has gained more importance because of the flexible nature of the cloud. In order to maintain security in cloud computing, the importance of developing an eminent intrusion detection system also increased. Researchers have already proposed intrusion detection schemes, but most of the traditional IDS are ineffective in detecting attacks. This can be attained by developing a new ML based algorithm for intrusion detection system for cloud. In the proposed methodology, a CIDS is incorporated that uses only selected features for the identification of the attack. The complex dataset will always make the observations difficult. Feature reduction plays a vital role in CIDS through time consumption. The current literature proposes a novel faster intelligent agent for data selection and feature reduction. The data selection agent selects only the data that promotes the attack. The selected data is passed through a feature reduction technique which reduces the features by deploying SVM and LR algorithms. The reduced features which in turn are subjected to the CIDS system. Thus, the overall time will be reduced to train the model. The performance of the system was evaluated with respect to accuracy and detection rate. Then, some existing IDS is analyzed based on these performance metrics, which in turn helps to predict the expected output. For analysis, UNSW-NB15 dataset is used which contains normal and abnormal data. The present work mainly ensures confidentiality and prevents unauthorized access.

**Keywords** Cloud intrusion detection system · Machine learning · Feature selection · UNSW-NB15 dataset · KLS framework

T. Sowmya (✉)
School of Engineering and Technology CHRIST (Deemed to be University), Bangalore, India
e-mail: Sowmya.t@res.christuniversity.in

G. Muneeswari
Department of Computer Science and Engineering, School of Engineering and Technology, CHRIST (Deemed to be University), Bangalore, India
e-mail: muneeswari.g@christuniversity.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_67

## 1 Introduction

Cloud computing is a rising technology that becomes an essential service in the entire IT world. Nowadays, because of the tremendous amount of dataflow in the network, it is necessary to maintain a storage system that should organize the data in a well-structured manner. The major benefits like scalability and pay has increased the popularity of cloud computing drastically. Although, security in cloud computing becomes a serious issue in this modern IT world because of the exchange of highly confidential information. In 2020, ransomware threat will contribute a serious hazard to the entire cyber world [1]. To overcome this challenge, many traditional systems have already been established and developed. Despite most of them were incapable of detecting the advanced type of attacks and opened the door for hackers. To complement these traditional systems, an intrusion detection system comes into the picture which can effectively detect intruders and prevent them from all the threats. Attacks can occur in the network as well as in the host system. There are two types of IDS based on network-based and host-based system [2]. In this digital world to compete with the intelligent attackers, many AI based IDS have already been proposed by integrating machine learning with IDS. Machine learning is a learning technology that learns from experiences and trains the model and makes some predictions based on the situations. While comparing with traditional intrusion detection mechanisms, the ML based IDS is more effective in detecting and reporting even unknown attacks to the network administrator [3]. Even though the underlying ML based IDS framework aims to detect the anomalies, the major drawback with these ML based IDS is to handle highly complicated datasets. Therefore, to increase the performance of these types of IDS based architecture, it is very important to perform feature engineering for highly complicated datasets. Therefore, to increase the performance of these types of IDS based architecture, it is very important to perform feature engineering for highly complicated datasets. This data monitoring and extracting features is a big challenge in this untrusted IT network. To overcome this problem, many machine learning-based techniques for extracting the features and reduction of features come into play. Feature reduction is an important step that comes under feature engineering which plays a vital role in machine learning. In feature engineering after extracting all the features, it is necessary to perform feature reduction to handle these huge volume data. The primary motivation of this paper is to propose an efficient ML based IDS incorporating an effective feature reduction technique to handle complex datasets.

The proposed method gives an expected output of high accuracy and highly effective detection rate comparing with the IDS technique without feature reduction. The work here focuses on contributing the following:

- An effective machine learning-based intrusion detection system for cloud is proposed which incorporates AI agent and KLS feature reduction techniques to handle complicated datasets.
- Fuzzy C-means clustering algorithm is followed for attack detection and random forest machine learning algorithm to classify the attack.

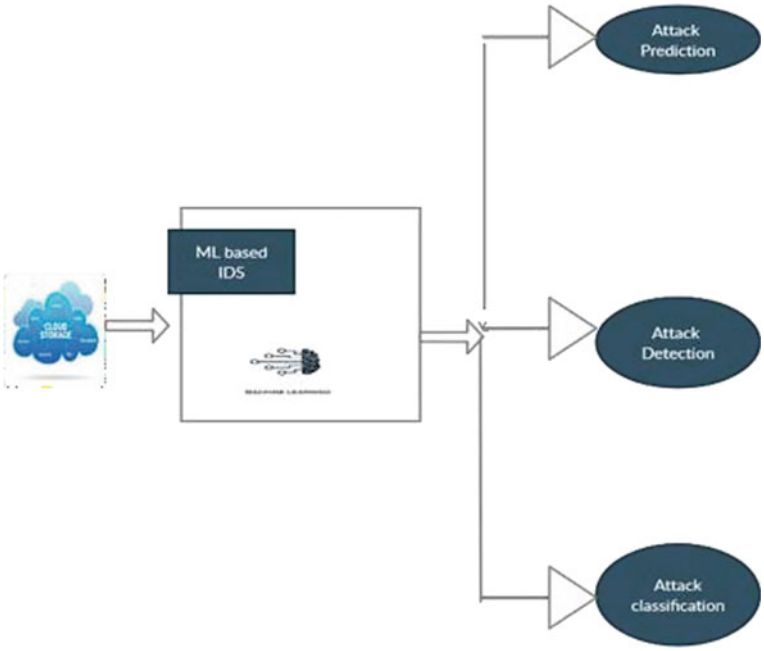



**Fig. 1** ML based IDS

- Our approach is to design a feature reduction framework that will select the best features to achieve better performance.
- Our proposed CIDS mechanism can detect all the attacks and protect the data from intruders.

Figure 1 explains the basic machine learning-based IDS that can detect, predict, and classify the type of attack. The primary motivation of this paper is to propose an efficient ML based IDS incorporating an effective feature reduction technique to handle complex datasets.

The paper is structured as follows: In Sect. 2, background of IDS and related work are introduced, and in third section, proposed methodology is discussed. Results and discussions are elaborated in Sects. 4 and 5 which covers the conclusion with further enhancements.

## 2 Background and Related Work

Since cloud security is a main issue for the entire IT world, cloud intrusion detection system monitors the traffic based on the data and selected features. In order to predict the anomalous behavior of a cloud, IDS monitors the traffic in a better way [4]. To

overcome the challenges present in the existing IDS, soft computing-based IDS is one of the measures to identify the intrusions in a better way. In [5], machine learning is a technology that models the system based on the experience. Machine learning-based IDS predicts the presence of intruders in the system by using different classification and regression algorithms.

The challenges faced by the existing IDS are given as follows:

(1) *Time complexity*
Every day along with the information technology world, data is also growing drastically. Hence, a mechanism is required to handle this data efficiently by selecting the best attributes from the entire attributes. So that it will train the data within a short period of time.

(2) *Space Complexity*
Due to this huge volume of data, interpretation of the output will increase the complexity. So in order to handle this huge volume of data, a feature reduction algorithm is necessary which will select only the relevant features.

(3) *Accuracy*
Accuracy can be achieved by reducing the volume of data; hence, an efficient feature selection method will increase the accuracy of the model.

(4) *Overfitting*
A powerful feature selection algorithm is needed to deal with the problem of overfitting. Complex data will always lead to overfitting, which gives an erroneous output.

Many researchers have already proposed ML based IDS to deal with all the challenges in cloud security. ML based IDS is also capable of ensuring security measures confidentiality, integrity, and availability.

Mukherjee and Sharma [6] propose an intrusion detection system with a feature reduction technique called feature vitality-based reduction method (FVRM) and naive Bayes algorithm for the classification of attack. Here, the author considers the performance criteria accuracy, true positive rate, and false positive rate; and investigation was conducted on NSL-KDD dataset. During the first phase of the implementation, FVRM based feature reduction is performed that reduces the features to 24, and for the next phase, intrusion detection is developed using naive Bayes classifier. Here, the author focuses mainly on DOS, Probe, R2L, and U2R attacks. These methods show an improvement in accuracy while comparing with CFS, IG, and GR feature reduction methods.

According to Nguyen et al. [7] an automatic feature selection method called correlation-based feature selection which will automatically select the features and reduces the complexity. They experimented using NSL-KDD benchmark dataset and investigated with best-fit CFS and genetic algorithm CFS.

In that, proposed method achieves higher accuracy and better performance while comparing with the above feature selection methods. In [8], an enhanced support vector decision function algorithm is proposed which is based on mainly two important steps. During the first step calculation of the weights from the functions takes place and for the next step ranking of the features takes place.

An approach based on a PCA combined [9] with random forest algorithm for classification of attacks reduces the complexity of the dataset. A dimensionality reduction technique called principle component analysis is used in the initialization step. Then, for the classification of the attacks, random forest algorithm is used. The proposed model shows better performance with respect to accuracy and detection rate.

Natesan and Balasubramanie [10] propose a multistage filter for network intrusion detection using random forest and PSO. Here, the method is composed of two stages, feature selection using PSO and classification of attack using random forest algorithm. Performance is evaluated with KDDCup 99 dataset, and the output is generated with high detection rate and low false alarm rates.

Hasan et al. [11], authors built two models using SVM and random forest for intrusion detection. They compared the performance metrics with respect to accuracy, precision, and false negative rate. In order to select the best intrusion detector, KDD CUP 99 is used. In [12], an feature selection algorithm is introduced which selects the features using filter and wrapper method with firefly algorithm. The selected features are subjected to C4.5 and Bayesian networks. For performance evaluation, KDD CUP99 dataset was used. The proposed model shows the highest accuracy while comparing with other models with the ten selected features.

## 3 Proposed Methodology

The work focuses on proposing an effective machine learning-based IDS in the cloud by using an intelligent agent-based KLS combinational framework for feature reduction and machine learning models for attack detection. This section begins with the data collection phase and later proceeds with the detailed step by step procedure of the system.

### *3.1 Data Collection*

In the proposed system, benchmark dataset UNSW-NB15 uses a labeled dataset characterized with 2 million records and 49 features for each record [13]. This dataset is publicly available for the researchers and established by UNSW cybersecurity laboratory records are classified as label ‘0’ for normal and ‘1’ for attacks. Attacks are further classified with nine latest attack types, namely Reconnaissance, Backdoor, Exploits, DoS, Analysis, Fuzzers, Worm, ShellCode, and at last Generic. The training and testing dataset has all the attacks and normal data (Table 1).

**Table 1** UNSWNB15 dataset classes

| No. | Type of attackers | Train UNSW-NB15 No. of record | Test UNSW-NB15 No. of record |
|---|---|---|---|
| 1. | Analysis | 2000 | 677 |
| 2. | Backdoor | 1746 | 583 |
| 3. | Dos | 12,264 | 4089 |
| 4. | Exploits | 33.393 | 11,132 |
| 5. | Fuzzers | 18,184 | 6062 |
| 6. | Generic | 40,000 | 18,871 |
| 7. | Reconnaissance | 10,491 | 3496 |
| 8. | ShellCode | 1133 | 378 |
| 9. | Worms | 130 | 44 |
| 10. | Normal | 56,000 | 37,000 |
| Total | | 175,341 | 82,332 |

## *3.2 Preprocessing of the Dataset*

Preprocessing is an important step in machine learning to increase the quality of the data since UNSW-NB15 is categorical data. In order to deal with this categorical data, one hot encoder is used for encoding the data during the first step. During the next step, in order to achieve better results, the Gaussian distribution standardization method is used to rescale the data.

## *3.3 Intelligent Agent Based Feature Reduction Method*

This section describes an intelligent agent system which uses two modules: data reduction agent and feature reduction agent. Feature reduction agent will select only the relevant features corresponding to the attack class. This research work uses UNSW-NB15 dataset for analysis. The proposed method expects to give better performance in terms of accuracy and detection rate. Here, the novelty introduced in the proposed architecture is the introduction of data selection agent which is a new approach which will filter the data that promotes the attack. The main objective of agent-based KLS framework is to reduce the complexity of UNSW-NB15 dataset. Here, this research work uses the intelligent agent-based KLS framework that integrates agent algorithm by using K-means clustering as the first stage and a combinational framework that combines SVM and linear regression algorithm in the second stage. The proposed model is expected to give a promising output in terms of feature reduction.

A strategy called an intelligent agent-based KLS framework is introduced to deduce the features from the entire subset of attributes. Even though UNSW-NB15 is already having classified data which corresponds to all the attacks, in a view to

increase the performance, the attack dataset is deployed to an unsupervised algorithm. In this section during the initial step, it uses K-means clustering algorithm, which clusters the training data into nine attack classes and a normal class. Since, K-means clustering is an unsupervised algorithm that can compute the centroids of each class and use them to join the training data from UNSW-NB15. Here, during the evaluation, the attack label should be dropped and normal from the training dataset. After partitioning the data into the attack and normal data using an intelligent data selection agent, the attack data is only filtered out and subjected as an input for the next step. For the next step, a combination of linear regression and SVM will help to extract only the features that provoke the attack and help to remove the irrelevant features

### 3.3.1 Intelligent Data Selection Agent

In our current literature, the agents work as an administrative tool which selects the attack data efficiently. Thus, during the feature reduction process, the novel data reduction agent will select only the data that provokes the attack.

**Algorithm1: Agent Based Data Selection Algorithm**
Input: A set of training data having 175,431 records from UNSW- NB15 dataset. Let it be' D'

Expected output: A set of reduced data's that contains only attack promoting data. Let it be 'R'

Intelligent data selection agent performs the following steps

*Step 1*: Initialize the cluster value with k=10

*Step 2*: Place centroids C for each cluster at random locations of D *Step 3*: for each centroid C, assign each data from D which is closer to C.

*Step 4*: for each centroid C, recompute C by calculating its mean value of all the data points.

*Step 5*: If no change in the centroid position, go to step 3

*Step 6*: Intelligent data selection creates 10 clusters with 9 attack data cluster [A1…A9] and 1 normal data cluster 'N'

*Step 7*: store the attack data cluster points [A1…A9] into R and output it.

Algorithm 1 explains the data selection process which employs K-means clustering algorithm which collects only the abnormal data in UNSW-NB15. After calculating the distance between the randomly selected centroids and data points, if the distance from the centroid 'C' and data points 'D' is minimum, then assigns the data point to the centroid 'C'. Repeat the calculation until there will be no reassigned data points. The purpose of this agent is to generate ten clusters with nine attack clusters and one normal data cluster. This allows the intrusion detection module to report the attacks easily. The present work aims to reduce the time complexity and to increase performance. When there is abnormal data, the data agent module expects to give a relatively efficient output.

### 3.3.2 Agent Based KLS Feature Reduction Method

UNSW-NB15 is having a total of 49 features, so it is very important to reduce the features to make the observations easy. Thus, a powerful feature reduction technique is proposed that can reduce the features only to relevant features. Here, the output from the previous step is taken as an input for this step. The selected attack data is passed as an input to linear regression and SVM algorithm. The two algorithms parallelly select the relevant features with the help of selected data from the previous step. During the final step, the output is generated by concatenation of the selected features from LR and SVM. The proposed KLS feature reduction method expects to give a high yield output in terms of accuracy and detection rate. The feasibility of the proposed intrusion detection system can be increased to a level higher than the earlier systems which are currently available. In addition to this, proposed model uses UNSW-NB15 dataset for evaluation which has all the latest attacks. Future researchers can use this model for feature reduction and can increase the performance of the system.

**Algorithm 2: KLS Algorithm**
Input: selected data 'R' and 49 features of the dataset Output: Set of reduced features 'F'

for each data and features of UNSW-NB15

(1) Extract the features from Linear regression algorithm, 'L'
(2) Extract the features from SVM algorithm, 'R'
(3) Concatenate the outputs of step 1 and 2, [L, R]
(4) Store the output of step 3 in' F'

In algorithm 2, a detailed overview of the agent-based KLS feature selection algorithm has been mentioned. The above-mentioned algorithm selects the best feature from the 49 features, and it has been expected to give a reduced number of features. Since SVM and linear regression are combined here, this strategy selects only the dominant features. It reduces the chance of overfitting, which will automatically lead to a highly accurate output. This agent-based method will increase the performance of our random forest classifier which in turn will increase the performance of the intrusion detection system.

## *3.4 Proposed Cloud Intrusion Detection System*

CIDS is a system that can monitor the traffic by detecting the attack and should report to the system administrator to prevent the attacks. Even though many traditional authorization mechanisms have been used, our cloud storage is still vulnerable to attacks. This section presents a CIDS framework to detect and classify the attack effectively. Here, KLS approach is used to build our IDS which accustoms the novel

data selection agent and feature selection agent. The CIDS framework is shown in Fig. 3 which comprised of two major steps, attack detection and attack classification. The attack detection phase detects whether an attack happens or not, and for the next phase, attack classifier detects the type of attack.

The CIDS application is used to achieve main functions feature reduction where KLS frame work is used for deducing the features as well as attack recognition. When structuring predictors, combining feature reduction methods can lead to high accuracy and detection rate by decreasing the false positive rate. Note that KLS is employing K-means, linear regression, and SVM. Data intelligent agent filters only the attack data, and the selected data is deployed as an input for the feature selection agent. The feature selection agent collects the selected data from the data selection agent, and based on the attack data, SVM, and LR together can reduce the features. By reducing the features, the complexity is reduced which in turn increase the accuracy and detection rate. Once the features are deduced, it will be used by the fuzzy C-means clustering algorithm to detect the attack in cloud by clustering process. When the data is transferred to fuzzy C-means clustering algorithm, an output is produced which indicates whether the data is normal or attack. Finally, if the output generated is attack, then the testing data's are transported to the random forest classifier, which classifies the attack based on the training data.

### 3.4.1 Attack Detection

After the agent-based feature reduction, the entire trained data is used to monitor the traffic. Here for distinguishing normal traffic and abnormal traffic, fuzzy C-means clustering algorithm is used. Fuzzy C-means algorithm divides the data into normal and attack. Even for the detection of unknown attacks, fuzzy C-means [14] clustering proved to be effective. As we are using agent-based feature reduction algorithm, performance expects to be significantly high.

### 3.4.2 Attack Classification

The efficient classification of attack is a major problem in the intrusion detection system. The next major phase of intrusion detection is to classify the attack using a major classifier, random forest. After detecting the attack, the trained data is transported through random forest classifier which is based on a voting method. Random forest along with agent-based feature reduction helps to develop effective intrusion detection by achieving high accuracy and detection rate than the existing algorithms.

The proposed CIDS framework (Fig. 3) combines the agent-based KLS feature selection algorithm using feature reduction agent (Fig. 2) to detect normal and abnormal data.

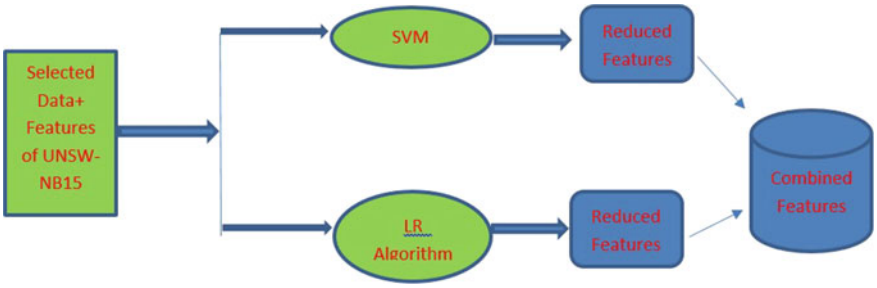



**Fig. 2** Feature reduction framework

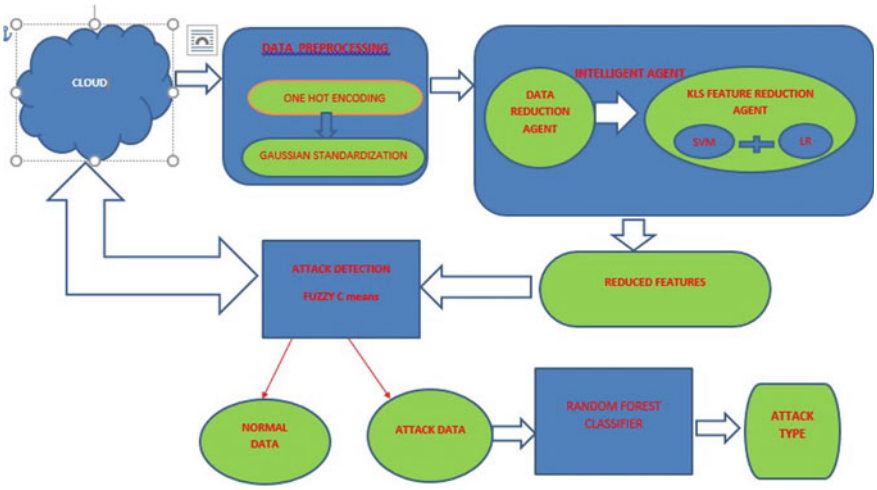



**Fig. 3** CIDS framework

## 4 Results and Discussions

To determine the performance of the intrusion detection system, UNSW-NB15 dataset has been used. This dataset is the latest dataset that has been widely used in intrusion detection strategy. Dataset has four major csv files, which is a combination of normal and attack data. An effective IDS should be capable of detecting all the attacks, and the performance metrics rate also should be high. The proposed agent-based IDS, which incorporates intelligent agents which use K-means clustering, linear regression, and SVM algorithm, expects to reduce the 41 features to relevant features. It helps to take the best features out of that 41 features by using the concatenation function in the KLS framework. For detection purposes, we are using fuzzy C-means clustering, which is an effective algorithm that detects the attack by clustering normal and abnormal data. Here, the classifier is random forest which classifies the type of attack by employing the majority voting method.

**Table 2** Performance analysis table

| Author name | Dataset used | Feature reduction method | Measurement analysis |
|---|---|---|---|
| Shone et al. [15] | NSL-KDD, K DD CUP 99 | Non-symmetric deep auto-encoder | Accuracy-98.81% |
| Xiao et al. [16] | KDD CUP 99 | PCA, AE | Accuracy-94.0% |
| Fangjun Kuang et al. [17] | KDD CUP | Genetics algorithm | Accuracy-99.69% |
| Li et al. [18] | KDD CUP | GFR | Accuracy-98.62% |
| Lin et al. [19] | KDD CUP | SVM | Accuracy-99.96% |

## *4.1 Performance Analysis*

To evaluate the accuracy and detection rate of machine learning algorithms, we used the intelligent agent mechanism for data selection and feature selection. It is expected that the present work shows a high detection rate and high accuracy as the combined approach of data selection and feature selection is employed. In addition to that for training purposes, we used the public dataset UNSWNB-15 dataset which in turn increases the accuracy and rate of detection.

### 4.1.1 Effect of Intelligent Agent

The results in Table 2 show the feature selection methods applied for intrusion detection. The following studies highlight how feature reduction methods affect the accuracy measure and detection rate. Even though researchers have already proposed eminent feature selection mechanisms, but still we need more investigations. Our proposal on UNSW-NB15 expects to reduce the features and recommends to implement a CIDS with high accuracy and detection rate.

The above performance analysis table focuses on different feature selection algorithms, and our current work expects to give promising results in terms of detection rate and accuracy. In an aim to increase the accuracy and detection rate, we used two-stage process for data selection. Our proposed model will show a very high accuracy rate comparing with the above-mentioned models. Since we are using an intelligent agent-based feature reduction mechanism, the time taken to model the IDS also will be less. In this current paper, we provided a solution to reduce the time taken to train the model with the reduced features. From Fig. 4 graph, it is clear that SVM shows the highest accuracy in feature reduction; hence, the proposed model is expected to show better accuracy.

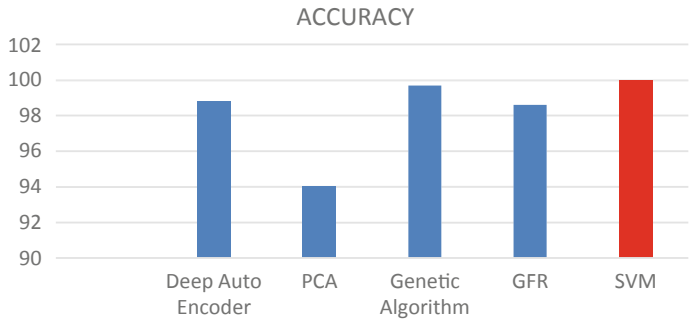



**Fig. 4** Performance comparison of feature reduction algorithms

## 5 Conclusion

The proposed research work used robust CIDS which uses an eminent feature reduction technique that will detect the attacks with high accuracy and high detection rate. Many traditional intrusion detection techniques are already available in the information technology world, our cloud data is still vulnerable to attacks, so it is necessary to provide security efficiently. The aim of the paper is to incorporate intelligent agent for feature reduction and the use of classifiers for the efficient intrusion detection system. From our current architecture, the proposed intelligent agent expects to be faster and can reduce the features. Randomly selecting the centroid data and computing the minimum distance between the data point and centroid makes the K-means clustering algorithm which makes the feature reduction process easy. Moreover combined approach of SVM and linear regression facilitates the feature reduction process by selecting the features corresponding to the selected data. Even for the detection of attacks, fuzzy C-means algorithm is used, and the robust random forest classifier classifies the attack efficiently. The above explanations provide clear justifications of why our proposed architecture expects to yield better results while comparing with the earlier methods. The current literature will serve as a base for further studies to develop an efficient cloud intrusion detection system (CIDS). Furthermore, the presented architecture can be implemented and can be extended for intrusion detection in the network. The proposed method on UNSW-NB15 dataset will give high performance by means of accuracy and detection rate. Further investigation is needed to enhance security in the cloud using machine learning and deep learning models.

## References

1. Raza M (2020) Top 5 cloud security trends of 2020. bmc July 28, 2020 Blogs https://www.bmc.com/blogs/cloud-security-trends
2. Aljamal I, Tekeoğlu A, Bekiroglu K, Sengupta S (2019) Hybrid intrusion detection system using machine learning techniques in cloud computing environments. In: IEEE 17th international

conference on software engineering research
3. Kwon D, Kim H, Kim J et al (2017) A survey of deep learning-based network anomaly detection. Cluster Comput 22:949–961. https://doi.org/10.1007/s10586-017-1117-8
4. Mehmood Y, Habiba U, Shibli MA, Masood R (2013) Intrusion detection system in cloud computing: challenges and opportunities. In: 2nd national conference on information assurance (NCIA), pp 59–66 [Online]. Available: https://doi.org/10.1109/NCIA.2013.6725325
5. da Costa KAP et al (2019) Internet of Things: a survey on machine learning-based intrusion detection approaches. In: Computer networks, Amsterdam, Elsevier Science Bv, vol 151, pp 147–157 [Online]. Available: http://hdl.handle.net/11449/185543
6. Mukherjee S, Sharma N (2012) Intrusion detection using Naive Bayes classifier with feature reduction. 4 [Online]. Available https://doi.org/10.1016/j.protcy.2012.05.017
7. Nguyen H, Franke K, Petrovic S (2010) Improving effectiveness of intrusion detection by correlation feature selection. In: International conference on availability, reliability and security, Krakow, pp 17–24
8. Sung AH, Mukkamala S (2004) The feature selection and intrusion detection problems. In: Advances in computer science—ASIAN. Higher-level decision making. ASIAN 2004. Lecture notes in computer science vol 3321. Springer, Berlin, Heidelberg [Online]. Available https://doi.org/10.1007/978-3-54030502-6_34
9. Waskle S, Parashar L, Singh U (2020) Intrusion detection system using PCA with random forest approach. In: International conference on electronics and sustainable communication systems (ICESC), Coimbatore, India, pp 803–808 [Online]. Available https://doi.org/10.1109/ICESC48915.2020.9155656
10. Natesan P, Balasubramanie P (2012) Multi stage filter using enhanced Adaboost for network intrusion detection. Int J Netw Secur Its Appl 4:121–135. https://doi.org/10.5121/ijnsa.2012.4308
11. Hasan Md Al, Nasser M, Pal B, Ahmad S (2014) Support vector machine and random forest modeling for intrusion detection system (IDS). J Intell Learn Syst Appl 6:45–52. https://doi.org/10.4236/jilsa.2014.61005
12. Selvakumar B, Muneeswaran K (2019) Firefly algorithm based feature selection for network intrusion detection. In: Computers & security, vol 81, pp 148–155. https://doi.org/10.1016/j.cose.2018.11.005.ISSN 0167-4048
13. Moustafa N, Slay J (2015) UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: Military communications and information systems conference (MilCIS), Canberra, ACT, 2015, pp 1–6 [Online]. Available https://doi.org/10.1109/MilCIS.2015.7348942
14. Ren W, Cao J, Wu X (2009) Application of network intrusion detection based on fuzzy C-means clustering algorithm. In Third international symposium on intelligent information technology application, Shanghai, pp 19–22 [Online]. Available https://doi.org/10.1109/IITA.2009.269
15. Shone N, Ngoc TN, Phai VD, Shi Q (2018) A deep learning approach to network intrusion detection. IEEE Trans Emerg Topics Comput Intell 2(1):41–50 [Online]. Available https://doi.org/10.1109/TETCI.2017.2772792
16. Xiao Y, Xing C, Zhang T, Zhao Z (2019) An intrusion detection model based on feature reduction and convolutional neural networks. In: IEEE Access 7 [Online]. Available https://doi.org/10.1109/ACCESS.2019.2904620
17. Kuang F, Xu W, Zhang S (2014) A novel hybrid KPCA and SVM with GA model for intrusion detection. Appl Soft Comput 18
18. Li Y, Xia J, Zhang S, Yan J, Ai X, Dai K (2012) An efficient intrusion detection system based on support vector machines and gradually feature removal method. In: Expert systems with applications, vol 39, Issue 1, pp 424–430. ISSN 0957-4174
19. Lin W-C, Ke S-W, Tsai C-F (2015) CANN: an intrusion detection system based on combining cluster centers and nearest neighbours. In: Knowledge-based systems, vol 78, pp 13–21. ISSN 0950-7051

# In-network Data Aggregation Techniques for Wireless Sensor Networks: A Survey


**T. Kiruthiga and N. Shanmugasundaram**

**Abstract** A sensor network consists of the random deployment of large numbers of tiny sized sensor nodes in a region of interest to detect the physical/environmental events and transmit the relevant data to the sink node through multihop communication. The nodes have many physical constraints like energy, memory, processing power, and hence the result is the limited or insufficient network lifetime of a network. To solve this problem, data gathering in an energy-efficient manner is an important task in the wireless sensor network to enhance network lifetime. Data aggregation is one such energy-efficient data gathering technique that reduces the data traffic and thereby the energy consumption substantially. The prime idea of the data aggregation is to gather, combine, and compress the data from various sensor nodes during transmission to the sink node. Among the available aggregation methods, in-network processing plays a major role to reduce the amount of data to be transmitted in the network. This article analyzes the various in-network data aggregation algorithms in detail and provides an insight into the techniques utilized.

**Keywords** Wireless sensor network · Data aggregation · In-network aggregation · Algorithms · Redundant data · Energy efficiency · Network lifetime · Aggregation rate

## 1 Introduction

In wireless sensor networks (WSNs), the sensor nodes are deployed in numerous applications. The sensor node performs the activities like sensing the information and communicating with one another and with the base station through the wireless network. To execute the behavior of all sensor nodes in the network is equipped with

T. Kiruthiga (✉)
Department of ECE, Vetri Vinayaha College of Engineering and Technology, Thottiam, Tamil Nadu, India

N. Shanmugasundaram
Department of ECE, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_68

the battery power for energy consumption [1]. Data aggregation plays an important role in combining the data from different sources which in turn is used as a powerful energy-saving mechanism and on the way after extracting the unnecessary data by reducing the energy consumption [2].

Based on the network structure, data aggregation is categorized into a tree-based aggregation technique, cluster-based aggregation technique, centralized aggregation technique, in-network aggregation technique, and hybrid structures. This article is prepared to furnish an organization of in-network aggregation by denoting the prime conception and overlaying the predominant and present specialization in the area of in-network data aggregation [3]. This article focuses on the comparative study of in-network aggregation of data in the networks. In-network aggregation is projected for resource constraint wireless sensor networks (WSNs) [4].

The essential scheme of in-network aggregation is that superfluous and inappropriate information is eliminated, and the significant information is merged into an aggregation consequence at intermediate nodes along the communication paths [5]. A communication cost is usually several times of level greater than the cost of processing the data [6]. Thus, the input range and congestion are decreased, which will conserve significant power [7]. This article intends to enhance the outlook of the in-network data aggregation and furnish an initiative and outset for the analyst who focused on these difficulties.

## 2 Data Aggregation

Data aggregation is the unique technique used in wireless sensor networks to extend its lifetime for saving energy by the nodes in communication. Data from different sources of nodes are aggregated at the intermediate from a source to sink based on some operation as shown in Fig. 1. Energy is wasted in the process of data gathering at a base station.

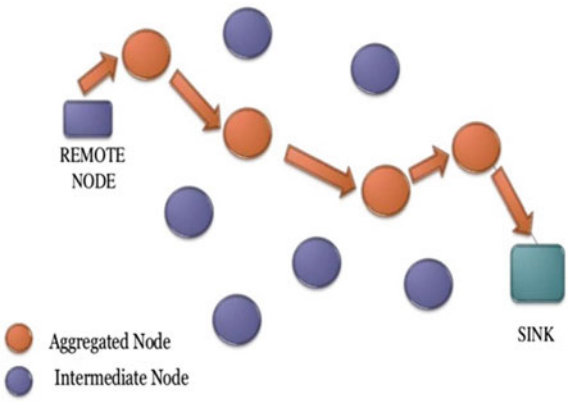



**Fig. 1** Data aggregation process

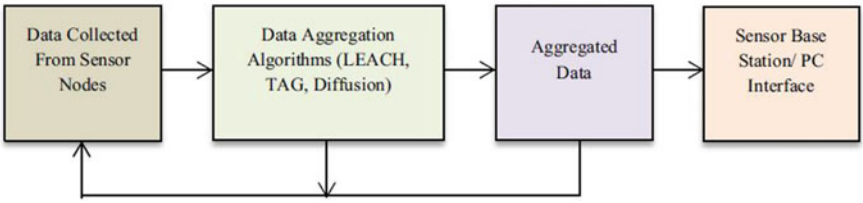



**Fig. 2** Block diagram of data aggregation

In the data aggregation method, instead of sending data out of all nodes to the sink node, a node called a data aggregator collects the information from its nearest nodes and aggregates them and finally transmits it to the sink node [1]. This method of aggregation includes the data packets transferred which contain only the needed information without redundant data [2]. When developing a new aggregation algorithm, few parameters to be thought-out namely the energy level of the sensor, resource constraint, and mathematical ability. The purpose of the data aggregation approach is to aggregate the sensed data in the node as shown in Fig. 2.

The different types of the algorithm such as hybrid energy-efficient distributed clustering approach (HEED), clustered diffusion with dynamic data aggregation (CLUDDA), power-efficient data gathering protocol for sensor information systems (PEGASIS), and a centralized approach is like to aggregate the observed input value from various sensor nodes in the network. A successful path from the aggregated data node to the sink node is selected for reliable transmission.

In the sensor networks, data aggregation consists of replacing each node of the individual sensing readings. For example, an aggregate function like MIN, MAX, AVERAGE (AVG), etc., can be used which allows the function of summarizing the n number of data obtained from all the intermediate nodes and send to the base station as only a single message from these n numbers of data [8].

This minimizes the excess of message forward to the base station and saves energy which is shown in Fig. 3. There are two classifications of data aggregation types namely data aggregation with range limiting and data aggregation without range limiting.

(A) **Data aggregation with range limiting**: This procedure combines and compresses the input value accepted from various motes to decrease the information delivered in the system. (e.g., for manipulating the overall n number of accepted data subsequently sends that information into a distinguished packet as an alternative for transmitting the n number of received data).

(B) **Data aggregation without range limiting**: This process compresses the received data from various nodes in a single packet without processing. (e.g., habit monitoring and disaster management cannot be processed simultaneously, but they can be sent in a distinguished packet which in succession to minimize the header).

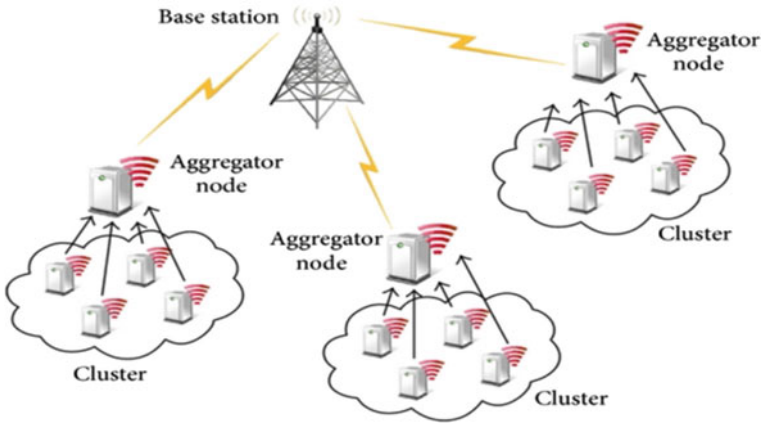



**Fig. 3** Data aggregations in sensor network

## 3 Taxonomy on Data Aggregation

The different data aggregation methodologies are grouped as a centralized aggregation technique, tree-based aggregation technique, cluster-based aggregation technique, and in-network aggregation technique are displayed in Fig. 4.

### *3.1 Centralized Aggregation Technique*

In this aggregation technique, one and all nodes in a network can convey the sensed data value to the mid node, which is considered as a strong node given power and bandwidth, etc., it is also called aggregator node. The purpose of the mid node is to aggregate the sensed data from various nodes in the network and then convey

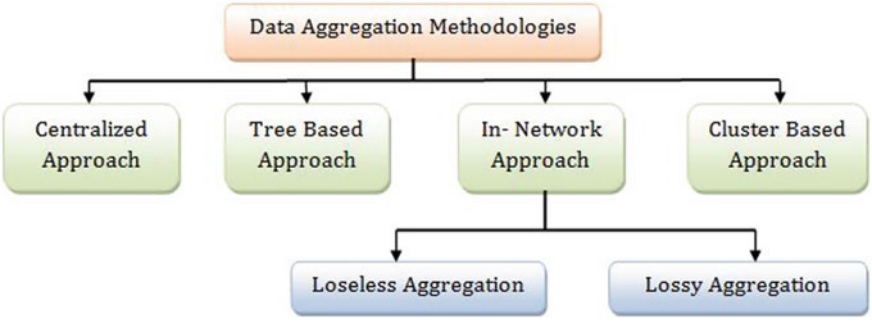



**Fig. 4** Organization of data aggregation

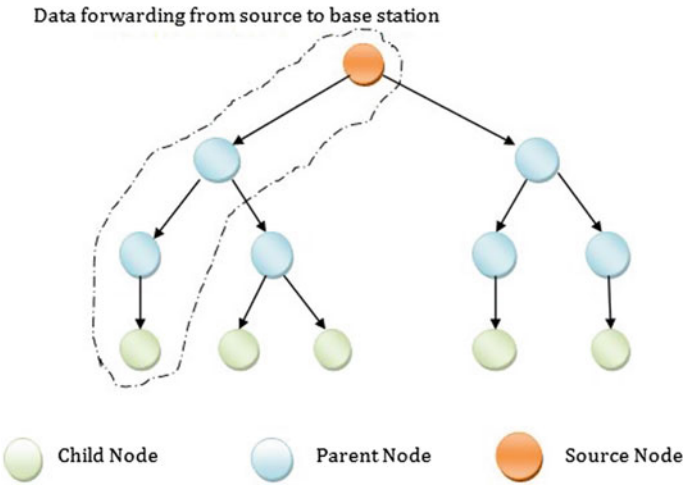



**Fig. 5** Structured tree technique

the information to the base station. This aggregation technique affects heavy traffic because of the huge number of data being transmitted [2].

### *3.2 Tree-Based Aggregation Technique*

In this aggregation technique, an aggregation tree is constructed at first with a minimum spanning tree. Here, root nodes act as a base station, terminal nodes act as a source node, and central nodes acts as parent nodes [2, 3, 9, 10]. Here, the leaf node transfers their observed data packets to their parent's node in a route found in the middle of the terminal node and base station is displayed in Fig. 5.

This aggregation technique has undergone some difficulties. In case, a packet loss occurring at any line of the tree will lose the entire data from the subtree, for the reason this approach needs a response mechanism to direct the aggregated data. Hence, the greedy incremental tree was recommended. It is formed by direct diffusion which initiates an effective path and grasping a new source to the initiated path. On noticing the first event, the data are directed using the shortest path for all new events [5].

### *3.3 Cluster-Based Aggregation Technique*

In this aggregation technique [2, 3, 5, 9, 10], the network is subdivided into several clusters. Within a cluster, a cluster head is selected, the process is to aggregate the data packet. Any other nodes in the network sense the data and transfer to the cluster

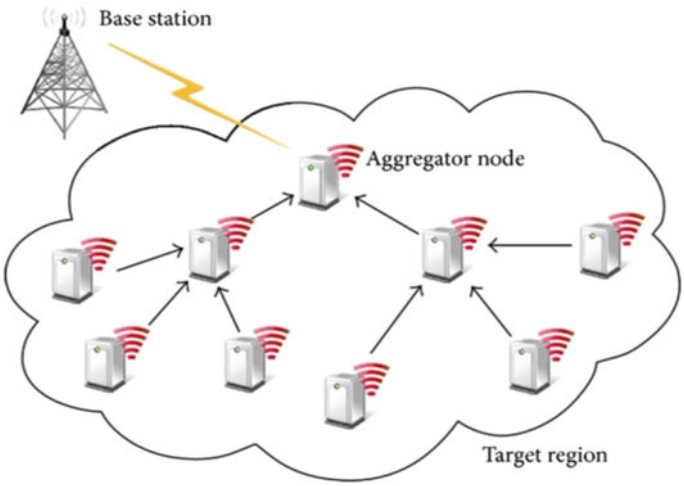



**Fig. 6** Cluster-based approach for data aggregation

head of the identical cluster in behalf of forwarding directly to the mid-station as shown in Fig. 6. This technique associates data fusion close to the specific nodes, and it can decrease the size of the data packet transferred to the sink. Hence, it saves the energy of the system [11].

### *3.4 In-network Aggregation Technique*

In this aggregation technique, sensed input values are aggregated at mid nodes to decrease the energy depletion. It also improves the network lifetime by decreasing the power depletion at each node [2]. Types of the proposal for in-network aggregations are:

(i) ***Lossy aggregation***: Data are sensed and collected from different nodes in the network and then a few aggregate tasks are involved on the sensed data such as SUM (), MIN (), MAX (), and AVG (). The range of the data packet is reduced. Individual purposive value is transferred to the base station instead of sending the whole packet of every node.

(ii) ***Lossless aggregation***: Lossy aggregation is necessary as it replies sensibly to the base station. Here, each packet is combined with a single packet without compression. For instance, in a jungle fire alarm, the lowest or highest thermal readings are needed promptly.

## 4 In-network Aggregation

In-network data aggregation permits the gradual processing of partial accumulation along with the separate way as for the routing tree to the root, thus the node is accountable to accumulate the entire outcomes of the aggregation. Such a thing decreases the amount of data packets to swapping in the middle of the adjacent nodes and therefore the energy depletion on the network is minimum and prolonging the duration of the node in the sensor network[12].

Usually, data aggregation is utilized to conserve power and energy efficiency. It is the action of arranging the various input values into one. WSNs may have redundant data because of more than one sensor; it can observe the same information when they are close together. Thus, the data aggregation eliminates unnecessary data in both transmission and reception.

In other cases, the number of communications can be reduced by aggregate queries. The queries are computed, based on the received data from every node in the network (e.g., if a query is asking for temperature in a particular region, then each sensor value should be averaged by aggregate function to the base station). Hence, there is no need for the base station to receive all the sensed values. Instead, the average value can be calculated within the network while the packets are transmitting via the base station. Aggregation is done by the intermediate nodes and then it routes to the base station; it is mention as in-network data aggregation.

In-network aggregation explores the circulation processing of data packets inside the network. It can be observed as a difficult process, because of this distributed network. Consequently, it needs cooperation between the nodes for excellent performance. The advantages of in-network aggregations are.

1. The number of packets that must be sent through the networks is decreased.
2. The likelihood of packet collisions is reduced.
3. The amount of redundancy received at the host nodes is decreased.
4. An improvement in the accuracy of results.

### *4.1 Routing Protocols*

A routing protocol is an important factor to be considered for in-network aggregation. When compared to classic routing, data aggregation requires various forwarding mechanisms [8, 13]. In a classic routing protocol, the data are forwarded on the short route to the destination. These data are accumulating to reduce the energy depletion level while the nodes should direct the packets based on the fulfillment and then it may select the next hop for in-network aggregation [3]. This way of data redirect is called data-centric routing.

## 5 Types of In-network Processing

In-network processing helps to minimize the size of data to be broadcast and helps to maintain the energy of the network. The techniques included for in-network processing are sensor fusion, sensor compression, sensor filtering, and sensor elimination [14].

### *5.1 Sensor Fusion*

Sensor fusion is a widespread technique which specific the aggregation of data as a partial action and its priority of information in place of data with the use of the approach such as signal processing, statistical analysis, machine learning, and probability. Sensor fusion processes all the sensed input values inside the network intending to get additional error-free data. To rise the value of the data collection, contributing the bandwidth, and expanding the network duration, sensor fusion is used [15]. Data fusion takes multiple sensed data and resulting sensed information is often outlined and consequently, the data are turned down.

### *5.2 Sensor Compression*

Squeezing is one of the most essential techniques to minimize data size. The power consumed will be more in executing the compression technique due to enormous data. Generally, the scheme of sensor compression includes the compression of the entire data packets from the sensor node before they are ready to transmit in the network and then the sink node will decompress the data packet while receiving it in the network [15].

This will end up with some energy depletion when compared to transferring the original data. Identifying the ratio of energy that makes use of both compression and transmission is the main issue. Usually, squeezing is the technique that generates the data close to the original data. The discrete cosine transform used in JPEG and MPEG is mathematically determined and it is problematic that energy-constrained sensors can operate it.

### *5.3 Sensor Filtering*

Data filtering is mentioned as the elimination of incorrect values or unknown values or noisy values. Data filtering is the part of data fusion that has been used to eliminate redundancy and noise. The goal of this approach is to duplicate the sensor value by

using methods such as Gaussian distribution and Bayesian-based approximation to handle the imprecision. It qualifies the future evaluation, elimination of prolongations and finally, it reduces the numerous broadcasts.

### *5.4 Sensor Elimination*

The way to eliminate the packets is to perform temporal aggregation or to decrease the data sampling rate of the sensor nodes. In temporal aggregation, a node caches the reading and compares it with the next sensing data. If there is not a major difference between the readings, it might not make sense to send it again. In this way, the data from the sensor can be reduced.

## 6 Algorithms of In-network Aggregation

In-network aggregation plays a major role in collecting, accumulating, and aggregating applicable data in WSNs. There are different in-network data aggregation classification techniques available which are shown in Fig. 7 and Tables 1 and 2 give out the comparison of those algorithms found on their common factor.

(A) ***DRINA*—**Data routing for in-network aggregation research is proposed by the subsequent authors. Thirumoorthy et al. [5, 13, 16–18] proposed an algorithm that reduces energy consumption and saves network lifetime by constructing

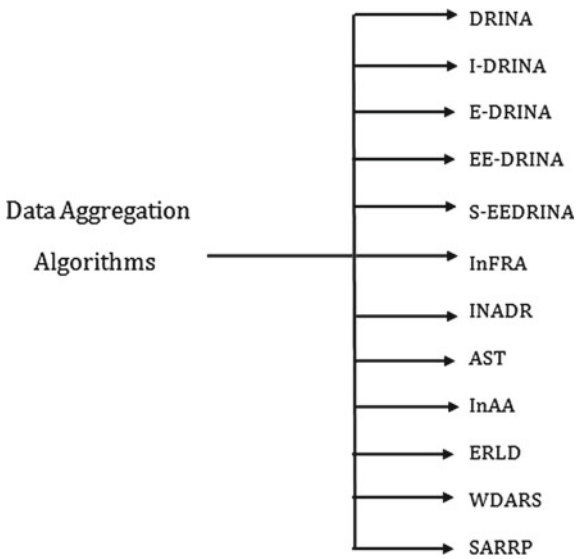



**Fig. 7** Various data aggregation algorithms

**Table 1** Summary of fundamental features of in-network aggregation

| Parameters/algorithm | DRINA | IDRINA | EDRINA | EEDRINA | SEEDRINA | InAA |
|---|---|---|---|---|---|---|
| Aggregation method | Cluster | Cluster | Cluster | Cluster | Cluster | Tree and graph |
| Energy consumption | Low | Medium | High | Medium | Medium | Low |
| Aggregation rate | – | Moderate | Min | Moderate | Moderate | Max |
| Network lifetime | High | Low | Low | Low | Low | High |
| Checking of alive nodes | No | No | No | No | Yes | No |
| Energy-saving mechanism | – | – | Periodic and timing strategy | Route redirection | – | – |
| Scalability | Medium | Medium | No | No | No | No |
| Query based | No | No | No | No | No | Yes |
| Route overlap | Max | Max | No | Max | No | No |
| Delivery rate | High | High | High | Moderate | Moderate | Low |
| Throughput | High | Moderate | High | Moderate | Moderate | Low |
| Security | Yes | No | No | No | Yes | No |
| Communication cost | Low | Low | Low | Low | Low | Low |
| Route repair mechanism | Yes | No | No | No | No | No |

the routing tree and increases the number of overlay routes and eliminates the redundancy data and thus it reduces the communication cost. DRINA will invariably increase data integration by the fault-tolerant method. It has four phases:

1. Hop tree construction
2. Cluster formation
3. Updating of hop trees
4. Route repair mechanism.

In the first phase, the hop tree is made by sending of hop configuration message (HCM) with initial value 1 from source to neighbor nodes. Each node verifies this value with the stored hop to tree value and updates it with the condition (i.e., stored value > received value) else it discards the message. In the second phase, aggregation points are selected. In the third phase, a new link is fixed by the coordinator with an establishment message to its next-hop node, and being updating node is reached. In the fourth phase, an acknowledgment ACK-based route repair mechanisms occur. When the node fails to get an ACK message, immediately it can select the new node, and thus the qualified link is formed.

**Table 2** Comparison of detailed parameters on in-network aggregation

| Parameters/algorithm | InFRA | AST | INADR | SARRP | ERLD | WDARS ara> |
|---|---|---|---|---|---|---|
| Aggregation method | Cluster | Tree | Cluster | Cluster | Cluster | Cluster |
| Energy consumption | Low | Low | Low | Medium | Medium | Low |
| Aggregation rate | Max | Max | Max | Moderate | Moderate | Max |
| Network lifetime | High | High | High | Low | Low | High |
| Checking of alive nodes | No | Yes | No | No | No | Yes |
| Energy-saving mechanism | Periodic scheme | – | – | – | – | – |
| Scalability | No | No | No | No | No | No |
| Query based | No | No | No | No | No | No |
| Route overlap | Max | No | Max | No | No | Max |
| Delivery rate | Moderate | Moderate | High | Moderate | Moderate | High |
| Throughput | Moderate | Moderate | Low | High | Low | High |
| Security | No | No | No | Yes | No | No |
| Communication cost | Low | Low | Low | Moderate | Low | Low |
| Route repair mechanism | No | No | Yes | No | No | No |

(B) ***IDRINA***—Ameya et al. [19] presented improved data routing for in-network aggregation (IDRINA), which is focused on in-network aggregation and data management in WSNs. This technique is a hybrid method of routing efficiency and energy efficiency and allows for trade-off communication complexity.

(C) ***EDRINA***—Mankirat Kaur et al. [9, 20–22] proposed enhanced data routing for in-network aggregation (EDRINA) is a cluster-based network. The primary goal of the algorithm is to increase network fulfillment and also recover the link failure problem. This algorithm makes use of the timing strategy for data aggregation and aggregates all the incoming packets into one packet as a result of aggregation.

In this algorithm, an intermediate node is attached outward on the clusters. The node in the cluster has an add-on battery backup when compared to the remaining nodes, subsequently the cluster node query about the battery backup of the pre-established link. The cluster head is adopted by the node which contains the highest priority of battery backup. At this moment, the adopted cluster head will take part in the connection and recover the link failure problem in the network.

This algorithm has three phases:

1. Setup phase
2. Initialization phase
3. Steady-state phase.

In the first phase, the task is initiated and performed by the sink. In the second phase, the cluster heads (CH) are selected from k nodes based on the probability of the node being CH. In the third phase, aggregated data are sent to the sink/base station by alive nodes within the network through CH.

(D) ***EEDRINA***—Shinde et al. [23] presented the energy-efficient data routing for in-network aggregation (EEDRINA) algorithm supported on route redirection. The route will increasingly meet to an alternate node disjoint the path with this following local redirection operation. Consider three successive nodes A, B, and C on the trail on the P-S connection. Node D is a neighbor node of all three nodes. Consequently transmitted data packets by three nodes can easily hear by node D. Node D can hear identical packets three times. By identifying this case, node D realizes that it can replace node B, therefore the subpath A-B-C are often redirected to A-D-C. Node D will do that process as long as the power level is quite that of B.

(E) ***SEEDRINA***—Sujatha et al. [24] proposed secure energy-efficient data routing for in-network aggregation (SEEDRINA) algorithm and offers safety for EEDRINA and has the accurate capacity to store the electricity to provide it for alive nodes. SEEDRINA has an eminent upgrade in average throughput and energy-efficient aggregation features.

(F) ***InFRA***—Nakamura et al. [25, 26] presented an information-fusion-based role assignment (InFRA) algorithm to classify the network by fixing roles to the nodes whenever the events are detected in the network. It initiates a hybrid group in which the source node is structured into clusters. The communication between the cluster and the sink node occurs only in a multihop manner. The primary goal of InFRA is to provide a logical estimation balance with tolerable service costs. The role assignment algorithms carry out the subsequent events:

1. Node detecting actions are conveyed to the collaborator and coordinator to form the cluster.
2. Routes are formed by connecting the clusters to the sink node where the relay is assigned to other nodes.
3. Service costs are decreased by information fusion.

(G) ***INADR***—Keerthishree et al. [10] proposed the in-network aggregated data routing (INADR) algorithm which is referred to improve the aggregation rate besides the relay pathway in stable through a fault-tolerant routing technique. This method has a minimum for setting up a routing tree. This algorithm aims to connect each node in the network to the sink node by creating the routing tree with the shortest path. When stably increasing the aggregation rate, the energy utilization will decrease in the network.

This method required only a minimum number of data packets from source to sink node and increases the number of overlay routes if there are excess events occur. High data aggregation is attained by stable data communication.

(H) ***AST***—Yi Zhang et al. [7] presented an adaptive spanning tree algorithm (AST), which can flexibly form and adapt an aggregation spanning tree. This algorithm has two phases:

1. Construction of a tree
2. Maintenance of the tree.

In this phase, one and all node sets a casual delay time to decide its father node via particular criteria. By this approach, it can ignore the good node which has more number of child nodes, and then it becomes congested. Asynchronously, a new node can replace any father node only by the use of the setting up a strategy. Constructions of the new tree with updating features are called AST.

***WDARS***—Mahdi et al. [27] proposed a weighted data aggregation routing strategy (WDARS) which focuses on increasing the overlay routes for effective data aggregation and connection rate problem in cluster-based networks synchronously. The proposed technique is analyzed for energy conservation, network lifespan, throughput, and packet delivery ratio.

The suggested protocol creates a scattered cluster and effective routing tree with the most energy saving and delay from overcrowding. When increasing the data aggregation, an event is detected, and the connection is established between all sensor nodes and the sink node.

The proposed algorithms contain three steps:

1. Hop tree establishment
2. Cluster formation and head selection
3. Route establishment by node weight.

In the first phase, a hop tree between the sensor node and sink nodes is organized. In the second phase, the formation of the cluster and cluster head selection starts when an event is sensed by the sensor node. In the third phase, events like path creation, data aggregation, and packet routing procedures occur.

***ERLD***—Du et al. [28] presented an efficient and real-time algorithm based on the dynamic message list (ERLD) algorithm which is formed by individual data array and dynamic list in which details of the data packets are stored.

The concept of ERDL has all cluster heads in the cluster equally react like filtering nodes. Rather than slow down the process of checking the redundancy in every transmission, a dynamic list is created and upgraded in all filtering nodes to save the record of data packets communicated by this node in certain intervals. The data packet is compared to all the records in the dynamic list when it reaches the filtering node. If the data packet is available in the list, then it will be rejected or else it will be communicated at once and the record of the list will be upgraded.

Based on the dynamic list, ERDL can be real time and conclude the message of whether imitated with the high ratio of filtering. There are three steps in ERDL:

1. Creating an initial list by setting filtering nodes
2. Transmitting messages after filtering
3. Dynamic updating list.

Firstly, a dynamic list is fixed by the filtering nodes which are employed in cluster heads. Then, the sensor receives the query packet and then the analysis is made to know the number of sensors included in this query and record the number of queries and send out all. At last, the range of the list would be logically powerful according to the list contained in the updated state.

(K) ***InAA***—Hua Yan et al. [29] proposed in-network aggregation algorithms (InAA), which can be used to discover the best aggregation node for various sensors from their fixed ancestor node in a particular multisensor query. To find the best aggregation node, it can also be used in both tree-based and graph-based structures. InAA contains three main procedures.

1. Establishing of sensor nodes in a stable form
2. Sending query packets to the sink node to the next hop
3. At the ancestor node data, packets are aggregated.

Initially, the sink node sends a stable level request packet with its level (0) to the neighbor. On receiving this, it will update its value (1) as a new level. Secondly, the sensor receives the query packet and then the analysis is made to know the number of sensors included in this query and record the number of queries and send out all. Finally, data packets get to the sink node. Concerning the query packets, if the level of the packet number is equal to the stored value, then it requires to direct the same packet. If the level of the packet number is lesser than the stored value, then it requires holding back the other data packets for aggregation.

(L) ***SARRP***—Anuradha et al. [30] presented the secure aggregated reliable routing protocol (SARRP) algorithm which is developed to combine clustering and aggregation techniques to build the best routing hierarchy with the most number of constructive lines that attach all the sensor nodes to the base station while making the best use of the available system resources in WSN. The SARRP algorithm can be divided into four modules.

1. Deployed nodes form special clusters using node energy levels and their neighborhood data.
2. Concentrates on data management and it provides exact data aggregated results for decreasing energy consumption and increasing throughput.
3. Focus on network connectivity.
4. Responsible for both setting up a reliable route among any sensor node to the base station and sink node with improved security features involving public key excess N cryptosystems.

## 7 Parameters for Analyzing In-network Aggregation

The effectiveness of the aggregation algorithms is assessed using the following parameters: quantity of alive nodes, throughput, energy efficiency, aggregation rate, network lifetime, scalability, security, packet delivery ratio is used in this survey are described below:

(A) ***Checking of Nodes Alive***: The capacity to sense and store the data in an exceedingly WSN depends on the group of alive nodes or there are the big qualities of nodes that could be operating well [24, 31]. At this moment, the effectiveness of the network is estimated by computing the variability of alive nodes within the network beyond a given duration. Checking of alive nodes in the aggregation process is used to define the network failure and storage occupancy per node per round [7].

(B) ***Average Throughput***: Average throughput is the ratio of variance among the large choices of data packets acquire at the destination node and the amount of value lost among the overall time required from the first data packet to the final records packet [24]. The throughput permits to regulate the perfect servicing of the device by estimating the efficient transmission during the periodic interval [32].

The purpose of throughput in data aggregation is to understand the data rates that are delivered to all the terminals in a network.

(C) ***Energy Efficiency***: Each sensor in a network consumes similar energy during the data collection but in a real situation, one and all nodes will ingest a different amount of power for data communication. Energy efficiency is defined as the percentage of the quantity of perfectly conveyed data in a network to the entire energy consumption for data to transmit. In data aggregation, energy efficiency is measured to enhance the network's lifetime.

$$\text{Energy efficiency} = \frac{\text{Quantity of successfully conveyed data}}{\text{Entire energy consumption to transmit those data}} * 100$$

(D) ***Network Lifetime***: The duration of the initial sensor node or collection of nodes in the network drain out of battery supply or the number of cycles gets disconnected due to defeat of more sensors.

(E) ***Data Aggregation Rate***: It is an action of fetching and gathering beneficial data value in a specific area. Data aggregation is a fundamental process to decrease energy consumption and to save resources. The aggregation rate is described as the percentage of the sum of aggregated data to the entire sum of sensed data. The main aim of this data aggregation rate is to denote the compressing capability of sensor nodes in the network during the aggregation process.

$$\text{Data aggregated data} = \frac{\text{Sum of aggregation data}}{\text{The entire sum of sensed data}} * 100$$

(F) ***Scalability***: Scalability is defined as the capability to hold a rising number of users in a network. A routing protocol is observed flexible concerning the network size.

(G) ***Security***: The security services in a WSN should protect the data communicated over the sensor network and save resources from attacks and misbehavior of nodes.

(H) ***Packet Delivery Ratio***: It is defined as the number of packets sent by the source node to the number of packets received by the destination node [24].

$$\text{Packet delivery ratio} = \frac{\text{Number of packets sent}}{\text{Number of packets received}} * 100$$

## 8 Comparison on In-network Data Aggregation Algorithms

Comparison of different in-network aggregation algorithms is a hard assignment because of the contrasting types of aggregation mechanisms and the lack of specifications. Hence, collation of the algorithm is important to realize the effectiveness of the in-network data aggregation technique which based on the fundamental characteristics of a parameter such as quantity of alive nodes, throughput, energy efficiency, aggregation rate, aggregation method, network lifetime, scalability, security, based on query approach, communication cost, energy-saving methods, packet delivery ratio, route overlap, and route repair mechanism are examined and briefly summarized in Tables 1 and 2.

As per the analysis in Tables 1 and 2, the main factors such as aggregation rate, energy consumption, and network lifetime are compared to different algorithms and the performance of the compared algorithms has been evaluated in Figs. 8 and 9.

The InAA, InFRA, AST, INDAR, WDARS algorithms recorded the highest data aggregation rate when compared to other algorithms. On the contrary, the algorithms like DRINA, InAA, InFRA, AST, INDAR, and WDARS consume less energy by in the view of energy nodes to steady the energy conservation between the nodes in the network.

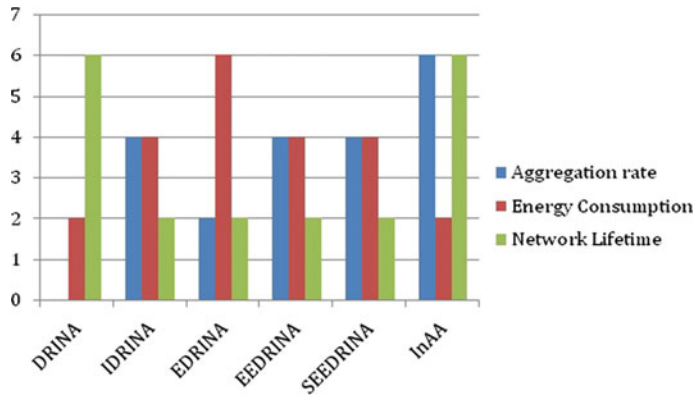



**Fig. 8** Comparison of DRINA-based algorithms

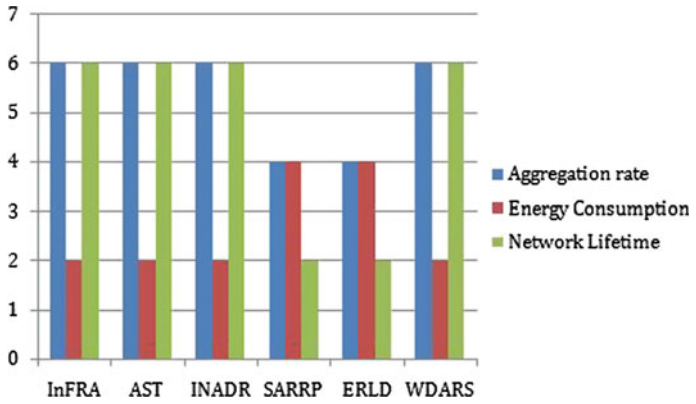



**Fig. 9** Comparison of various parameters on a different algorithm

Next, on considering the network lifetime the algorithms like DRINA, InAA, InFRA, AST, INDAR, WDARS show the best result out of all the remaining algorithms.

## 9 Conclusion

Data aggregation is one of the predominant techniques used for reducing power conservation and enhancing network lifespan in wireless sensor networks (WSNs). The motive behind in-network aggregation is to reduce the redundant data during sensing and thereby reducing the traffic in data transmission. This review article focuses on analyzing various in-network data aggregation algorithms for their efficiency based on important parameters like energy efficiency, aggregation rate, and network lifetime. Other factors such as the number of alive nodes, throughput, energy-saving techniques, scalability, network security, route overlap, and route repair mechanisms are also taken into consideration for analysis and summarized. This article will give a deeper insight into data aggregation based on in-network processing and motivate the researchers to develop more efficient data aggregation algorithms in the future.

## References

1. Randhawa S, Jain S (2017) Data aggregation in wireless sensor networks-previous research, current status and future directions: wireless communication. Springer, Berlin. https://doi.org/10.1007/11277-017-4674-5
2. Shilpa SG, Meenakshi Sundaram S (2017) Data aggregation techniques over wireless sensor networks—a review. Proc Int J Eng Res Technol (IJERT'17) 5(22)

3. Fasolo E, Rossi M (2007) In-network aggregation techniques for wireless sensor networks: a survey. IEEE Wirel Commun
4. Raja M, Data R (2018) Efficient aggregation technique for data privacy in wireless sensor networks. IET Netw. https://doi.org/10.1049/iet-net.2017.0104
5. Joshi GM, Patil BM (2016) Data routing in-network aggregation for wireless sensor network. Proc Int J Comput Appl (IJCA'16) 137(3)
6. Lu Y, Kuonen P, Hirsbrunner B et al (2016) Benefits of data aggregation on energy consumption in wireless sensor networks. IET Commun 6(14):2189–2197
7. Zhang Y, Pu J, Liu X, Chen Z (2014) An adaptive spanning tree-based data collection scheme in wireless sensor networks. Proc Int J Distrib Sens Netw (Hindawi)
8. Ennajari H, Maissa YB, Mouline S (2017) Energy efficient in-network aggregation algorithms in wireless sensor networks—a survey. Springer, Berlin. https://doi.org/10.1007/978-981-10-1627_11
9. Shrivastava N, Kawitkar R (2014) EDRINA for more battery life in Wireless sensor networks. In: Proceedings of international conference for convergence of technology-(ICCT'14). https://doi.org/10.1109/i2ct.2014.7092158
10. Keerthishree PV, Vani KS (2014) An efficient method for reliable routing in-network aggregation in wireless sensor network. Proc Int J Eng Res Technol (IJERT'14) 3(4)
11. Arora V, Sharma TP (2016) In network aggregation techniques and data management in wireless sensor networks: a survey. Proc Int J Adv Comput Eng Netw 4(7)
12. Dai X, Xia F, Xia F, Wang Z, Sun Y (2006) An energy-efficient In-network aggregation query algorithm for wireless sensor networks. In: Proceedings of the first international conference on innovative computing, information and control, 7695–2616
13. Bijjal SS, Shivamoorthy RC (2014) Survey on data routing for in-network aggregation: a lightweight and reliable routing approach for in-network aggregation in wireless sensor networks. Proc Int J Eng Res Technol (IJERT'14) 3(27)
14. Ari I, Akkaya K (2011) In-network data aggregation in wireless sensor networks. Handb Comput Netw. https://doi.org/10.1002/9781118256114.ch70
15. Chen Y, Sheng J, Zhang S, Liu L, Sun L (2009) Data fusion in wireless sensor networks. In: Proceedings of the second international symposium on electronic commerce and security, IEEE Computer Society. https://doi.org/10.1109/ISECS
16. Thirumoorty P, Karthikeyan NK, Sudha S, Manimegalai B (2014) A review on routing protocols for in-network aggregation in wireless sensor networks. Proc Int J Innov Res Sci Eng Technol (IJIRSET'14) 3
17. Nakade V, Chavhan N (2015) Implementation of DRINA algorithm for energy efficient routing in wireless sensor network. Proc Int J Sci Res Develop (IJSRD) 3(2):441
18. Kaur M, Kaur B (2015) Review paper on DRINA protocol. Proc Int J Adv Res Comput Commun Eng (IJARCCE) 4(9)
19. Bhatlavande AS, Phatak AA (2015) Energy efficient approach for in-network aggregation in wireless sensor networks. Proc Int J Current Eng Technol 5(4)
20. Rajasekaran R (2014) An efficient data-centric routing approach for wireless sensor networks using EDRINA. Proc Int J Appl Innov Eng Manag (IJAIEM'14) 3(3)
21. Manni R, Rai MK, Kansal L (2014) EDRINA: enhanced data routing for in-network aggregation algorithm for WSNs. Proc Asian J Inf Technol 14:276–280, 18:250–260
22. Kaur M, Sharma A, Kaur B (2016) A novel technique for link recovery in energy efficient DRINA protocol for Wireless sensor network. Proc Int J Adv Res Comput Commun Eng (IJARCCE'16) 5(1)
23. Shinde YY, Sonavane SS (215) An energy-efficient critical event monitoring routing method for wireless sensor networks. Proc Int J Comput Appl (IJCA'15) 114:10
24. Sujatha B, Jilo CT, Roa CS (2018) Energy efficient data route in-network aggregation with secure EEDRINA. Lecture notes on data engineering and communications technologies, pp 1–9. https://doi.org/10.1007/978-981-10-6319-0_1
25. Nakamura EF, Loureiro AAF (2008) Information fusion in wireless sensor networks. In: Proceedings of ACM SIGMOD international conference on management of data—SIGMOD'08. https://doi.org/10.1145/1376616.1376775

26. Nakamura EF, Ramos HS, Vilas LA et al (2009) A reactive role assignment for data routing in event-based wireless sensor networks. Comput Netw 53 (Springer)
27. Adil Mahdi O, Abdul Wahab AW, Idris MYI, Znaid AA (2016) WDARS: a weighted data aggregation routing strategy with minimum link cost in event-driven WSNs. Proc J Sens (Hindawi)
28. Du T, Qu S, Liu K, Xu J et al (2016) An efficient data aggregation algorithm for WSN based on dynamic message list. In: Proceedings of 7th international conference on ambient systems, networks and technologies, computer networks 83:98–106 (Elsevier)
29. Yan H, Al-Hoqani N, Hua S (2018) In-network multi sensors query aggregation algorithm for wireless sensor networks database. IEEE Internet of Things
30. Anuradha MP, Ramya S, Doraipandian M (2018) SARRP: secure aggregated reliable routing protocol for wireless sensor networks. Proc Int J Pure Appl Math 118(20)
31. Bongale AM, Nirmala CR, Bongale AM (2020) Energy-efficient intra-cluster data aggregation technique for wireless sensor network. Springer, Berlin, Proce Int J Inf Technol. https://doi.org/10.1007/41870-020-00419-7)
32. Jennifer S, Raj AB (2019) QOS optimization of energy efficient routing in IOT wireless sensor networks. Proc J ISMAC 1:12–23

# Comparative Analysis of Traffic and Congestion in Software-Defined Networks


**Anil Singh Parihar, Kunal Sinha, Paramvir Singh, and Sameer Cherwoo**

**Abstract** The different methods used for classifying traffic along with the prediction of congestion and performance in software-defined networks were discussed. Although congestion prediction has foreseen many challenges, the algorithms did not give very accurate results. But over a period of time, several methods have been discovered to identify and predict the performance and congestion in software-defined networks (SDN). In this article, various techniques of classification were compared and predicted through tables and graphs.

**Keywords** Software-defined networks · Traffic classification · Performance prediction · Machine learning · Congestion

## 1 Introduction

The evolution of the Internet has made it easier for people to access anything from anywhere in the world. This has led to a new era of digitalization or Digital India. However, despite being used and accepted everywhere, traditional Internet networks (IP-based) are pretty complex and very hard to manage. With the discovery of new communication techniques like 4G upcoming 5G, it has become really important for the carrier service providers and operators to manage a high volume of load and data over their limited capacity network bandwidth since more and more users are getting added to the list and usage is increasing day by day [1].

So companies cannot afford to provide poor quality streaming services to their customers, and hence, there is a necessity at the moment to deal with the problem of over congestion of the network.

Software-defined networking is comparatively a new concept, which helps to fix the shortcomings of conventional networks which include storage needs and various environments of scalable computing. It is dynamic, financially savvy, versatile, and ideal for the high data transmission and dynamic nature of the present organization

A. S. Parihar (✉) · K. Sinha · P. Singh · S. Cherwoo
Department of Computer Science and Engineering, Delhi Technological University, Delhi, New Delhi, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_69

applications. These particular abilities make it deployable in many service network environments, from home and venture organizations to servers in cloud organizations.

## *1.1 Concept of SDN*

The SDN architecture consists of different layers that describe the way the network devices in the data plane interact with the controller through a southbound API. The API provides an interface for interaction between the planes. For this purpose, there are many protocols available, but the OpenFlow is preferable in the project. The control plane of SDN consists of controllers like POX and OpenDaylight. The northbound API is an interface between the control and application layer (management plane). The statistics about the data plane such as the flows are gathered through REST API [2] (Fig. 1).

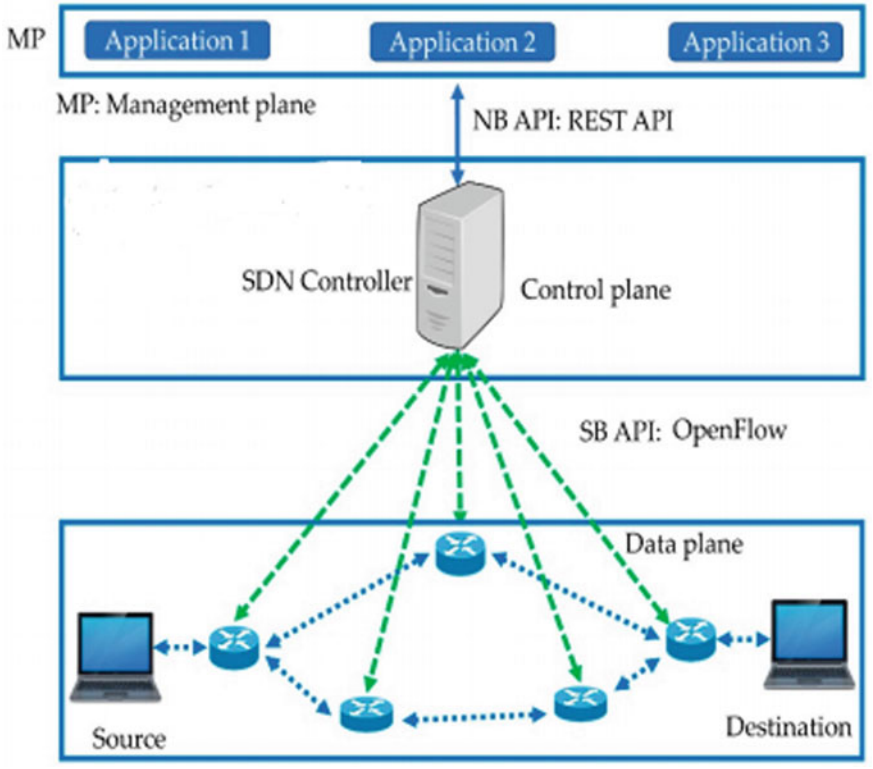



**Fig. 1** Architecture of SDN

### *1.2 SDN Over Traditional Networks*

There is an enormous recovery delay in traditional IP networks inferable from the flooding of packets, the expanded time for congestion prediction, calculation of substitute ways, and updating the routing table. Nonetheless, in SDN, the controller has global control of the network. Along these lines, it chooses ideally while scanning a proficient substitute way for the disrupted flows. Also, the controller screens the end-to-end connectivity, and accordingly, when a network is congested, the controller can reconfigure the network to restore the end-to-end availability for all paths. As opposed to conventional networks where each node floods it with packets to discover a substitute way, the SDN furnishes the arrangements with less intricacy and adaptability. The programmability and adaptability can be utilized to dynamically apply arrangements in the network through the control plane, according to the changing QoS prerequisites when the connection disconnects. Consequently, the time, cost, and workforce are decreased.

The major problems faced by the present network systems are fault issues, security issues, and congestion issues [3–5]. However, network congestion has always been an area of concern for networks since it reduces the quality of service for the users. This situation usually arises when a link between two nodes has a capacity (bandwidth) less than the data load which it is carrying. The congestion of the flow table is also a major problem. There are some approaches to solve it by compressing the table [6] when a network is congested, it hampers the QoS of a network and users face difficulties and poor quality of services when using Internet services. Software-defined networks help to reduce this problem of congestion. Congestion can be avoided by either limiting the rate or by early prediction. Thus, to predict link congestion in SDN, various machine learning techniques have been used earlier which aim to improve the accuracy and prediction rate of congestion and performance of a network [7].

## 2 Areas of Study

### *2.1 Traffic Classification*

The software-defined networking mechanisms of the OpenFlow protocols provide an excellent platform to implement network control applications that are based on machine learning. One of the major studies in this area includes understanding and classifying the type of data that can be gathered. There is an increase in the volume of data flowing through networks. With an increase in the complexity of the applications which generate this data, assembling information from this generated data for interpreting and predicting its impact has become of prime importance to perform efficient network administration. There have been many types of IP classification done previously: IP-based, Port-based, ML-based, etc. SDN data has also been used for ML-based classification [8] using SDN-based traffic gives many advantages as

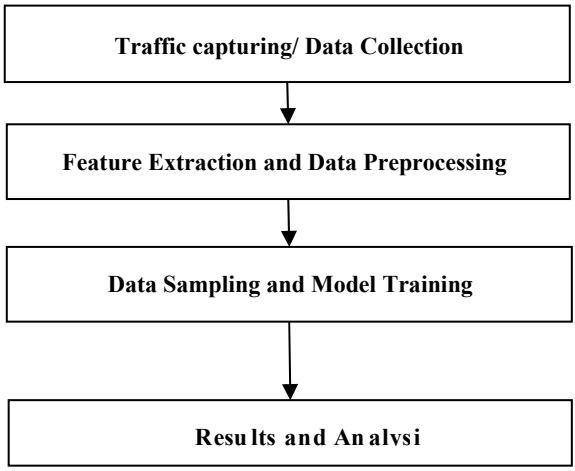



**Fig. 2** Process of traffic classification

the statistics are gathered directly at the controller so it allows to directly use that information where the control plane is also present. The controllers can directly manipulate the flow rules based on the classification of data. Previous work includes a collection of various types of traffic packets from different types of applications such as BitTorrent, Dropbox, and YouTube, and various classifiers have been used like stochastic gradient boosting, random forests, and extreme gradient boosting [9] (Fig. 2).

The primary kind of data which needs to be classified is collected via simulating traffic in SDN. Then certain parameters are chosen as features. After this preprocessing is done in which we normalize our data. Later, the data is divided into subsets of training, testing, and validation. Training data is used for our model. Then the model can be applied to the controller to obtain the desired results.

## *2.2 Congestion Prediction*

One of the major problems in networking is the congestion of a link due to the increase in demand for network resources. This leads to an increase in delay of packet transmission and rate of packet loss which in turn adversely affects quality of experience (QoE). A lot of previous work has been done in this field. Most of them are TCP-based while a few of them are UDP-based. The UDP generally has higher throughput with faster transmission time and less delay; therefore, it has a wide variety of applications. The parameters selected as eigenvalues were receiving and sending rates of a switch, bandwidth of the link, switch load, and end-to-end source-to-destination traffic rate [3] (Fig. 3).

Initially, the data that needs to be classified is collected via simulating traffic in SDN. Congestion factor (CF) is calculated based on features from our data collected.

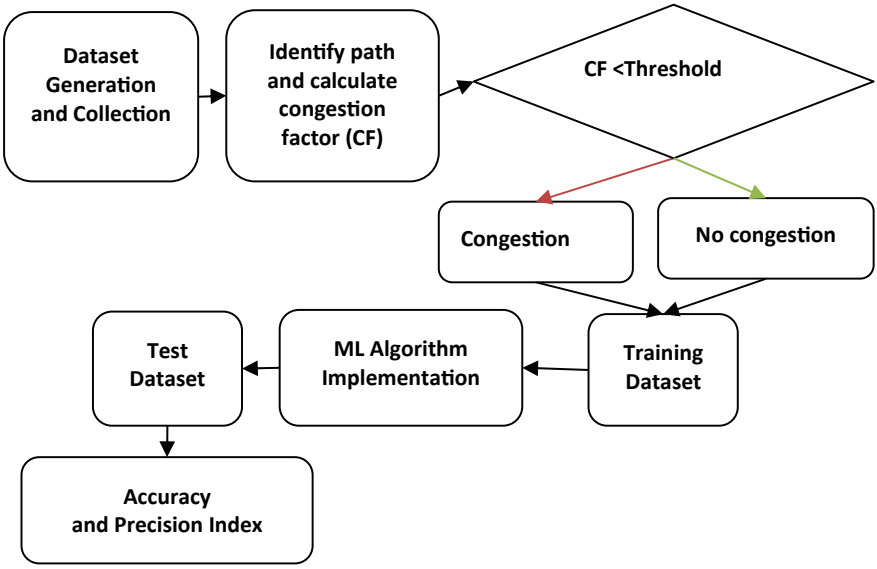



**Fig. 3** Process of congestion prediction

If the CF is a greater threshold value, it is classified as congested. The same logic along with extracted features is used to train our model and apply the same to our controller to take necessary action whenever there is congestion in our network.

### 2.2.1 Congestion Metrics

In traditional networks, there is an uneven distribution of traffic and load. Hence, there should be some threshold value to evaluate the possibility of a link being congested. Normally, a high load on a link greatly impacts the network's stability and performance.

To measure the link bandwidth utilization, the following formula has been used

$$\mathrm{BU}_{\mathrm{Link}} = \frac{\mathrm{Bits}_{\mathrm{Sent}} + \mathrm{Bits}_{\mathrm{Recieved}}}{\mathrm{Time} * \mathrm{Bandwidth}_{\mathrm{Link}}} \tag{1}$$

Whenever the value of $\mathrm{BU}_{\mathrm{link}}$ exceeds 0.7, then the link is said to be congested (refer to Table 1). By combining it with the conventional description of the congestion-degree criteria by experts and the operators, it can be visualized that 70% is selected as the critical value of bandwidth link utilization to determine the degree of congestion of the link. When it exceeds 70%, traffic scheduling strategies and corresponding congestion control strategies must be taken up.

While studying traffic metrics and using the same for the dataset, it is essential to choose the type of data transfer that will take place and corresponding parameters that

**Table 1** Standard congestion definition

| Standard network link congestion level | Average utilization of link bandwidth |
|---|---|
| Severe congestion | More than 90% |
| General congestion | 80–89% |
| High load | 70–79% |
| Normal load | Less than 70% |

will be collected during that transfer. For traffic classification, TCP data transmission has been used, and for traffic congestion prediction, UDP data transmission has been used in the dataset for effective analysis [10].

## 3 Algorithmic Analysis

### *3.1 Machine Learning*

It forms a small but essential part of AI. It is a powerful technique of training machines so that they can improve from their experience and learn: all this, without being programmed explicitly. The machine can determine from the data accessible in its environment (or experience). This is further used to enhance the overall performance. It is further classified into two types.

- Supervised learning methods learn from the data in the past and it generalizes for future data. The data in supervised learning is a 'labelled' data. For example, let us say that a training dataset consists of *X*-*Y* pairs, (*X* being the input and *Y* being the output) where the machine learns an algorithm that evaluates an appropriate output for sample input.
- Unsupervised learning is used to draw surmisings from datasets consisting of information without labels. The most popular unaided approach for learning is analysing clusters (K means), used for identification of hidden patterns among given data points, done by grouping similar data together in a single cluster.

### *3.2 Neural Networks*

A neural network is an aggregation of smaller units called neurons. It is inspired by the working of the human brain. It forms the base of deep learning. These networks take in data, train themselves to recognize the patterns in this data, and then predict the output for a new set of similar data. Some attributes of neural nets include self-organization, adaptive learning, and fault tolerance. They are useful in clustering, classification, and pattern recognition [11].

**Table 2** Methodology for dataset collection [3, 9, 13]

| Congestion prediction | Traffic classification | Performance prediction |
|---|---|---|
| 13 switch, 12 hosts. Multiple pairs of source and destination hosts are used to transmit data. Network data is collected every second in real time with each experiment lasting 12 h. The total amount of data is 559929 | Two distinct datasets were used, one consisting of an unlabelled set of traffic information, and the second dataset (smaller in size) was a labelled one and contained the applications which generated the data | 5 OpenFlow switches. Each is connected to two hosts. POX controller is connected to all switches and hosts. SCP is used to send files of different sizes between two hosts. Runtime was divided into three parts with 60 min each. File size transferred was 100 MB, 50 MB, and 25 MB, respectively. |

# 4 Methodology

## *4.1 SDN Simulation*

In most of the studies conducted, Mininet was used to simulate the network topology and collect the data. OpenFlow switches are connected along with some hosts. Controllers like POX, OpenDaylight are also used in the topology. Data is sent among different hosts for a certain period and is collected at different intervals of time. Different types of performance metrics are collected for different areas of study. For example, flow table, RTT, and throughput are collected in performance prediction. In link congestion prediction, iperf was used to simulate UDP traffic, and five parameters were collected. In traffic classification, the controller determines the protocol by analysing the packet. For example, TCP flags are examined to determine the TCP protocol [12] (Table 2).

## *4.2 Application of Machine Learning*

The collected dataset due to the result of the above SDN simulation then is used to train our machine learning or deep learning model. In the case of performance prediction, the dataset was split up into three parts [13]

- Training set
- Validation set
- Testing set.

After collecting and preprocessing the dataset, it was split randomly into 70%, 15%, and 15%, respectively, for the above-mentioned points. This was then used to train the artificial neural network. For link congestion prediction, four models were trained: 1DCNN, MLP, KNN, and SVM. [14] For traffic classification, accuracy for three classifiers was calculated: stochastic gradient boosting, random forests, and

extreme gradient boosting. The best accuracy model is determined and then applied to the controller to make intelligent decisions and consequently manipulating the flow tables for the switches in the topology.

## 5 Results and Analysis

### 5.1 Traffic Classification

Following were the results obtained after the implementation of the algorithms. Accuracy is determined by taking the ratio of correct predictions over the total number of predictions as described in (2).

$$\text{Accuracy} = \frac{T_{+ve} + T_{-ve}}{T_{+ve} + F_{+ve} + T_{-ve} + F_{-ve}} \tag{2}$$

where $T_{+\text{ve}}$ is true positive, $T_{\text{-ve}}$ is a true negative, $F_{+\text{ve}}$ is false positive, and $F_{-\text{ve}}$ is a false negative. The results are promising with great accuracy values for all the websites from which the traffic dataset was collected [15] (Fig. 4).

### 5.2 Congestion Prediction

Following were the results obtained after the implementation of all algorithms. The graph below shows the classifiers accuracy and precision scores [16].

For SVM, the obtain values by ranging the value of penalty parameter C, between 5 and 15, The used kernel functions are linear, polynomial, radial basis function, and sigmoid, which is shown in (3)–(6).

$$\text{KF}(s_i, s_j) = s_i^T s_j \tag{3}$$

$$\text{KF}(s_i, s_j) = \left(\gamma s_i^T s_j + r\right)^d, \gamma > 0 \tag{4}$$

$$\text{KF}(s_i, s_j) = e^{\left(-\gamma ||s_i - s_j||^2\right)}, \gamma > 0 \tag{5}$$

$$\text{KF}(s_i, s_j) = \tanh\left(\gamma s_i^T s_j + r\right) \tag{6}$$

For MLP, the input data is a $1 \times 5$ feature vector. There are two hidden layers, whose units vary from 30 to 70 for performance comparison [14].

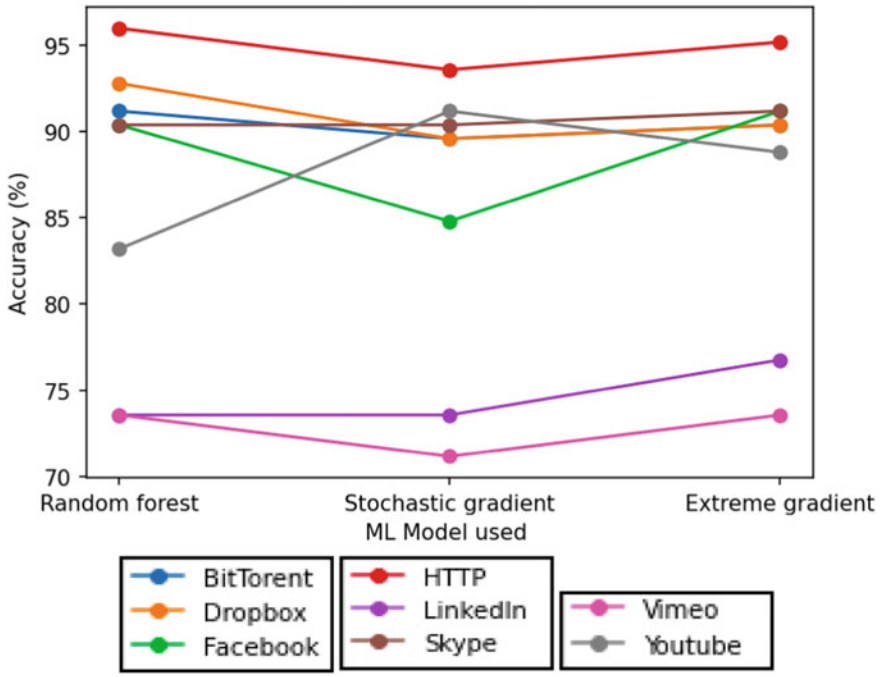



**Fig. 4** Comparison of classifier accuracy for the three models

In KNN, Euclidean distance has been used to find the distance between pairs of data points. The parameter k is varied from 5 to 15 for analysis.

Parameters for CNN include batch size = 200, learning rate = 0.05, epochs = 40, input size = 1 × 5 array. Several convolution layers can be around 5, convolution filters vary from 20 to 36, and each convolution layer is followed by a pooling layer and set the down sampling factor to 2. At last, a fully connected layer was produced.

Precision is the fraction of the number of positive class predictions that actually belong to the positive class which is shown in (7).

$$\text{Precision} = \frac{T_{+ve}}{T_{+ve} + F_{+ve}} \tag{7}$$

where $T_{+\text{ve}}$ is true positive and $F_{+\text{ve}}$ is false positive (Fig. 5; Table 3).

## 6 Conclusion and Future Scope

This article compares the different methods used for the classification of SDN traffic using machine learning. It also describes a variety of machine learning and deep

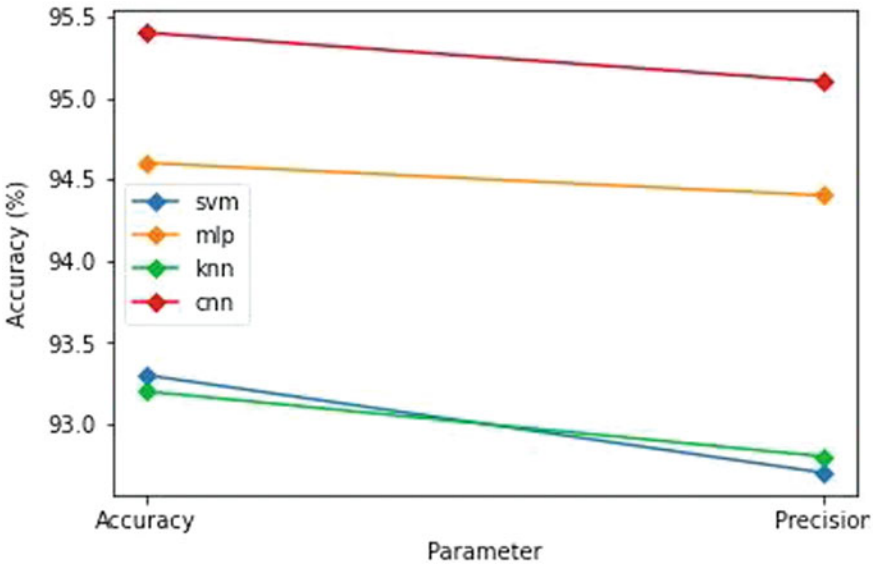



**Fig. 5** Comparison of accuracy and precision for different models

**Table 3** Comparison of algorithms

| ML model | Accuracy | Precision |
|---|---|---|
| SVM—Support vector machine | 0.933 | 0.927 |
| KNN—K-nearest neighbours | 0.932 | 0.928 |
| MLP—Multi-layer perceptron | 0.946 | 0.944 |
| 1DCNN—Convolutional neural networks | 0.954 | 0.951 |

learning-based techniques used for predicting link congestion in software-defined networks. In classification, XGBoost was found to show the best classifier accuracy. In congestion prediction, it was seen one-dimensional convolutional neural networks showed the best accuracy from the four used algorithms.

In future, a larger and denser dataset with a greater variety of parameters was utilized. For future works, our aim is to use a larger and denser dataset with a greater variety of parameters. We aim to use a larger topology network (more than 15 nodes) for SDN simulation. Our major focus would be the usage of deep learning, more specifically the usage of recurrent neural networks for prediction purposes. The aim is to apply the model to the controller and predict the congestion rate along with the best path that it should follow so that the controller can redirect the traffic through the least congested path.

## References


1. Saied W, Souayeh NBYB, Saadaoui A, Bouhoula A (2019) Deep and automated SDN data plane analysis. In: IEEE international conference on software, telecommunications and computer networks, Croatia
2. Smys S, Raj JS (2019) A stochastic mobıle data traffic model for vehicular ad hoc networks. J Ubiquitous Comput Commun Technol 1:55–63
3. Wu J, Peng Y, Song M, Cui M, Zhang L (2019) Link congestion prediction using machine learning for software-defined-network data plane. In IEEE international conference on computer information and telecommunication systems (CITS)
4. McGregor A, Hall M, Lorier P, Brunskill J (2004) Flow clustering using machine learning techniques. In: Proceedings of the 5th international passive and active network measurement international workshop, PAM, France
5. Azzouni A, Boutaba R, Pujolle G (2017) NeuRoute: predictive dynamic routing for software-defined networks. In International conference on network and service management (CNSM), Tokyo
6. Leng B, Huang L, Qiao C, Xu H (2016) A decision-tree-based on-line flow table compressing method in software defined networks. In IEEE/ACM 24th international symposium quality of service (IWQoS), Beijing
7. Azzouni A, Pujolle G (2018) NeuTM: a neural network-based framework for traffic matrix prediction in SDN. In: NOMS 2018-2018 IEEE/IFIP network operations and management symposium, Taipei
8. Fan Z, Liu R (2017) Investigation of machine learning based network traffic classification. In: International symposium on wireless communication systems (ISWCS), Bologna
9. Amaral P, Dinis J, Pinto P, Bernardo L, Tavares J, Mamede HS (2016) Machine learning in software defined networks: data collection and traffic classification. In: IEEE 24th international conference on network protocols (ICNP), Singapore
10. Xiao K, Mao S, Tugnait JK (2019) TCP-Drinc: smart congestion control based on deep reinforcement learning. IEEE Access 7:11892–11904. https://doi.org/10.1109/access.2019.2892046
11. Sendra S, Rego A, Lloret J, Jimenez JM, Romero O (2017) Including artificial intelligence in a routing protocol using Software Defined Networks. In: IEEE international conference on communications workshops (ICC Workshops). https://doi.org/10.1109/iccw.2017.7962735
12. Latah M, Toker L (2018) Artificial intelligence enabled software defined networking: a comprehensive overview. IET Netw. https://doi.org/10.1049/ietnet.2018.5082
13. Sabbeh A, Al-Dunainawi Y, Al-Raweshidy HS, Abbod MF (2016) Performance prediction of software defined network using an artificial neural network. In: SAI computing conference (SAI). https://doi.org/10.1109/sai.2016.7555965
14. Nikravesh AY, Ajila SA, Lung C-H, Ding W (2016) Mobile network traffic prediction using MLP, MLPWD, and SVM. In: IEEE international congress on big data (BigData Congress). https://doi.org/10.1109/bigdatacongress.2016.63
15. Powers DMW (2011) Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. Int J Mach Learn Technol 2(1):37–63. https://arxiv.org/abs/2010.16061
16. Kumar S, Bansal G, Shekhawat VS (2020) A machine learning approach for traffic flow provisioning in software defined networks.in: International conference on information networking (ICOIN), Barcelona

# A Comparative Analysis on Sensor-Based Human Activity Recognition Using Various Deep Learning Techniques


**V. Indumathi and S. Prabakeran**

**Abstract** To forecast conditions of action or actions during physical activity, the issue of classifying body gestures and reactions is referred to as human activity recognition (HAR). As the main technique to determine the range of motion, speed, velocity, and magnetic field orientation during these physical exercises, inertial measurement units (IMUs) prevail. Inertial sensors on the body can be used to produce signals tracking body motion and vital signs that can develop models efficiently and identify physical activity correctly. Extreme gradient boosting, multilayer perceptron, convolutional neural network, and long short-term memory network methods are contrasted in this paper to distinguish human behaviors on the HEALTH datasets. The efficiency of machine learning models is often compared to studies that better fit the multisensory fusion analysis paradigm. The experimental findings of this article on the MHEALTH dataset are strongly promising and reliably outperform current baseline models, comprising of 12 physical activities obtained from four separate inertial sensors. The best efficiency metrics were obtained by MLP and XGBoost with accuracy (92.85%, 90.97%), precision (94.66%, 92.09%), recall (91.59%, 89.99%), and F1-score (92.7%, 90.78%), respectively.

**Keywords** Human activity recognition · Sensors · IoT · LSTM · CNN · MLP · XGBoost

## 1 Introduction

Continuous growth in the health sector has led to astronomical advancements in the field of medicine. Due to this continuous growth, the quality of life has greatly increased when compared to one hundred years ago. Everything from life expectancy,

V. Indumathi
Department of Computer Science and Engineering, R.M.K Engineering College, Kavaraipettai, Tamil Nadu, India

S. Prabakeran (✉)
Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_70

physical health, education, safety, and freedom has vastly improved. This rise in health sector growth has led to an increase in healthcare costs. Steadily increasing medical expenses have led to dramatic cost-cutting steps employed by healthcare providers worldwide in tracking patients with chronic disease, monitoring the aged, along with many other cases. New technologies in healthcare will certainly lead to lowering the cost of health care by ensuring that physicians, surgeons, and other medical workers work and perform their everyday tasks in the vicinity of the hospital more effectively.

Choosing the right sensor to match the working environment or application is critical when faced with a human activity recognition challenge. With advanced sensors created each year, the number of different types of sensors to choose from is infinite. Continuous research in the area of behavior identification has driven more firms to try to reap the advantages of forecasting events to boost coordination and efficiency. The numerous types of tasks undertaken for behavior identification are discussed in this section. Body-worn inertial sensors monitor physical exercise activities like riding, jogging, standing still, and powerwalking in related jobs. When these tasks are completed, they generate a specific type of range of body motion, with calculated accelerations that are relatively identical when done by people with different characteristics.

Recognition of human behavior used with powerful technology will theoretically benefit from remote patient control, elderly people, chronic condition patients, and living with environmental assistance. Simple activities such as cycling, running, and jogging have been successfully recognized and classified to date. Complex activities are proving increasingly difficult to monitor, with continuous active research conducted in this area of HAR. The main goal of HAR is to predict common activities in real-life surroundings. Researchers are exploring pattern recognition and human–computer relationships due to their applicability in the real world, such as a human activity recognition healthcare framework. Successfully classifying human activities through wearable sensors generates endless individual information, which provides insight into the individuals' functional ability, lifestyle, and health. In this research, the MHEALTH dataset is analyzed using a variety of deep learning models. These models aim to classify activities performed by volunteers based on data gathered from on-body inertial sensors. The exploratory analysis distinguishes the differences and similarities between these deep learning models throughout this research.

The overall aim is to identify which algorithm best suits the data while discovering which algorithm best classifies each body movement of each person based on vital signs recordings. Two data clustering algorithms analysis identifies relationships between feature attributes and pleasantly visualize the data. Using wearable sensors, human activity recognition (HAR) involves recognizing the physical movements of a subject by analyzing data produced from on-body wearable sensors. Accelerometers, gyroscopes, and magnetometers are these inertial sensors, while the movements are known as everyday living activities (ADL). As described in [1], ADL involves self and body and emphasizes mobility specifically. Due to the applicability of sensor fusion, sensor-based HAR dominates the current study, which requires the incorporation of sensor data from multiple sensors, which drives analytical results in terms of

reliability, accuracy, and completeness. Deep learning methods continue to continuously advance and strengthen the HAR area in this regard. With its in-depth expertise and analytical capacity, XGBoost leads the way in taking data-oriented classification tasks and selecting and processing invaluable features from the data effectively. There are four deep learning (DL) models, which are applied to the HAR problem in this article. Using on-body sensor signal data created from four different sensors, several models were developed, trained, and analyzed for the results to identify which model best fits the data in terms of precision, accuracy, recall, F-score, and the total number of misclassified instances.

This article demonstrates its ability to perform parallel optimization and tree pruning while restricting overfitting and constantly learning sparse features, XGBoost is the highest performing model. Section 2 provides a summary of the work on the recognition of human behavior. The remainder of this paper is structured the following. Section 3 offers a description of the MHEALTH dataset, the design, and the research methodology. Section 4 addresses experiment performance. Finally, in Sect. 5, the conclusion and the future scope of the research work were presented.

## 2 Related Work

Recognition of human behavior using wearable or mobile sensors supports a variety of applications such as health care, exercise, smart home, etc. For example, medical teams may track elderly people's health conditions based on information about their activity. Daily energy expenditure was estimated and provided with good advice according to the level of operation of the users. [2]. The growth of smartphones and the increase of access to technologies like the availability of high-speed internet and network infrastructure have significantly changed the lives of people. Today, several smartphones contain a range of powerful sensors, including orientation, location, network, and direction sensors. Specifically, motion or inertial sensors (e.g., accelerometers) were commonly used to detect the physical movements of the users. [2]. The sensor is a device that senses and obtains the changes that occurred in the environment and redirects the collected information to the operating system. Smartphone sensors are categorized into three major groups called motion, environmental, and position sensors. Motion sensors use axis-based sensing approach to finding the measurement. Environmental parameters are measured by environmental sensors, for example, temperature, humidity, and light. Position sensors are used for measuring the distance of the reference position. Widely used smartphone sensors are,[3]

1. Accelerometer: It detects variations in smartphone orientation concerning the *x*, *y*, and *z*-axis.
2. Ambient light sensor: It senses the light density of the environment. Auto-brightness adjustment in the mobile phone is the best example of this type of sensor.

3. Barometer sensor: Atmospheric pressure is sensed by the barometer, this sensor assists the GPS to track the location in an efficient manner.
4. Gyroscope sensor: It finds the axis-based motion along with angular rotation so the clean data can be obtained.

Human activity recognition can be achieved through a video-based or sensor-based approach. Video-based HAR investigates videos or photographs that include human movements, while sensor-based HAR focuses on movement data from smart sensors such as an accelerometer gyroscope, Bluetooth, sound sensors, etc. HAR approaches are generalized into a certain type of body-worn sensor, object sensor, ambient sensor, and hybrid sensor. Physical activities that are performed by users can be detected by using body-worn or ambient sensors embedded in smartphones. The physical activities of the users are directly related to the movement and resting of the human body. The body-worn sensors, such as the accelerometer, magnetometer, and gyroscope, are the sensors that users may wear. The environmental changes can be identified by ambient sensors. There are few such ambient temperature sensors, radars, motion sensors, and sound sensors [4]. Human activities are categorized into seven groups such as ambulation, transportation, daily activities, exercise, kitchen activities, transitional activities, and self-care activities. Based on the category, activities are shown below [5]

(a) Ambulation: Sitting, standing, running, lying, falling
(b) Transportation: Driving a car, riding a bicycle.
(c) Daily activities: Watching TV, drinking, eating, using a phone, using a computer, reading the book, listening to music, sleeping.
(d) Transitional activities: Walking upstairs and downstairs, lying down and getting up, sitting down and getting up.
(e) Self-care activities: Combing hair, shaving, brushing teeth, washing hands, washing face, washing clothes, drying hair.
(f) Kitchen activities: Adding tea-bag, add sugar, add milk, removing tea-bag, pour milk, making coffee and tea, cooking pasta, cooking rice, feed fish.

Wesllen et al. stated the essential steps required to recognize the human activity: Data collection, segmentation, feature extraction, and activity classification.

1. Data collection: Extraction of raw data from different sensors embedded in a smartphone. The data must be adequate to produce good models for classification activities. To ensure the correctness of the activity model, some parameters must be considered like sampling frequency, the position of the smartphone, and orientation from the user, and data collection time [6]. But Foerster et al. [7] stated that the technique pursued to collect the raw data from the user is very difficult for any human activities. The accuracy level of ambulation activities for controlled data collection is 95.6% but the accuracy level is dropped to 66% for the natural environment. The proper analysis would recognize a significant number of users with different characteristics. This will give a better result for the new user without obtaining extra training data.

2. Segmentation: The raw data or inertial sensor signals are not recommended to take a decision in the classification process. The raw data, therefore, requires other transformations, such as breaking the continuous raw sensor data into the windows over a certain period. Noise removal from the signal is another important role of segmentation. Inertial sensor signal might have a noise which leads to misclassification; obviously, it affects the model accuracy. Using signal processing techniques like low-pass, high-pass, and Kalman filters, the noisy data can be removed easily [6].
3. Features: Settings defined in the segmentation stage played a vital role to extract the sensor features. The extraction algorithm takes the input from time windows. The selection of good features is a very important factor to classify the labels correctly. For any classification model, the accuracy can be directly affected because of low-quality features. Wesllen et al. described a new principal domain. It is divided into three groups such as time, frequency, and discrete domain. Time-domain uses mathematical approach statistical data are extracted from the signals. Repetitive patterns are captured through frequency domain. The discrete domain makes the signal pattern by converting sensor signals into symbols.

Thus, although the main domain permits for the chaining of features, it is very important to remember that all of them are orientation-dependent features when the time features are handled alone, and when these features are chained with the magnitude or vertical–horizontal components, they all become orientation-independent features. Some of the time domain features are min, max, amplitude, amplitude peak, sum, absolute sum, Euclidian norm, mean, absolute mean, mean square, mean, cross-validation, auto-correlation, skewness, kurtosis. Example of frequency domain: Energy, energy normalized, power, centroid, entropy, domain component [6].

Bashar discussed the deep neural networks, and the model accuracy is expelled even human performance. This survey on the deep learning neural network architectures utilized in various applications for having an accurate classification with an automated feature extraction especially in CNN [8]. Prabhakaran et al. describe the various clustering framework for predicting kidney disease [9]. Indumathi et al. explain the utilization of various machine learning models [10]. This project details a deep learning comparative study of the:

(a) Convolutional neural network model.
(b) Long short-term memory (recurrent neural network) model.
(c) Extreme gradient boosting (XGBoost) model.
(d) Multilayer perceptron.

## *2.1 Convolutional Neural Networks in HAR:*

Jiang and Yin [11] compare multiple deep convolutional neural network (DCNN) architectures using accelerometer and gyroscope data in classifying activities. Jiang et al. [11] perform analysis on the UCI MHEALTH dataset (UCI) [12], USC-SIPI

human activity dataset (USC) [13], and a dataset compiled by the fusion of smartphone motion sensors (SHO) [14]. Jiang et al. [11] compared performances to identify the architecture, which achieved the highest accuracy, recall, and precision along with the low computational cost. Jiang et al. [11] found that SHO achieved the highest accuracy, followed by USC, while UCI performing slightly lower. Hammerla, Halloran, and Plotz [15] compare multiple convolutional and recurrent approaches when using wearable sensors in classifying activities. Hammerla et al. [15] perform analysis on three datasets: The opportunity dataset, Pamap2 dataset, and Daphneit Gair (DG) dataset. Hammerla et al. [15] conduct thousands of experiments to identify the substantial effect of altering hyperparameters. Performance evaluation indicated that the approaches achieved the highest accuracy on DG, the lowest root mean squared error, and the highest F1-score, followed by Opportunity, while performance scores on Pamap2 were slightly lower. Kim and Moon [16] compare the use of deep convolutional neural networks (DCNNs) for activity recognition in classifying activities. Kim et al. [16] also compare DCNNs for human detection. A Doppler radar gathers data, which produces velocity data when placed on a human or near a human. Kim et al. [16] found that the DCNN achieved accuracy as high as 97.6% for human detection. They also found that human activity classification accuracy reached heights of up to 90.9%.

## *2.2 Long Short-Term Memory (LSTM) in HAR*

LSTMs are intended on tackling the vanishing gradient problem. The main difference between LSTMs and RNNs is LSTM's use of memory cells. Memory cells allow for the sufficient storage and sequential processing of data. Time is not restricted and the data does not disappear back into the network. It enables the development of relationships in the data, leading to insightful knowledge regarding the output to be analyzed. Gating is at the core of LSTMs. Gating regarding LSTMs involves component-wise multiplication of the input as seen in related work [17]. This leads to consistent updates in each data cell, due to the gating calculation that applies to each cell. The data must encounter the write, read, and reset gates to process data correctly. The write gate is the input gate. The read gate is the output gate while the reset gate is the forget gate. LSTMs contain information in a gated cell, which is the key idea of these networks. LSTMs can add or delete information to the cell through the gates. These gates are composed of a sigmoid neural network layer and a point-wise multiplicative operator.

## *2.3 Extreme Gradient Boosting in HAR*

Ayumi investigates if extreme gradient boosting is superior in classifying activities in the HAR domain over classical techniques such as support vector machine (SVM) and

Naïve Bayes (NB). The UT Kinect-Action3D Dataset, the Badminton Sports Action Dataset, and the Bali Dance Motion Dataset conduct analysis. XGBoost takes more computational time to run as opposed to the other two methods but prevails as the best method with higher accuracy, precision, recall, and F1-score. Zhang et al. [18] propose an XGBoost method to recognize activities on their own dataset, which they created themselves. The dataset consists of 40 volunteers performing multiple activities contained in an indoor facility. XGBoost outperforms other ensemble classifiers with a higher recognition rate in accuracy, F1-score, and precision. F-score reached heights of up to 84.41% while accuracy surpassed previous studies achieving a rate of 84.19%. Nguyen, Fernandez, Nguyen, and Bagheri [19] explained, the XGBoost model uses wrist-worn accelerometer data to identify events, RGB-camera data, and environmental sensor data. In contrast to previous research, this unique method produced an elevated performance of 38% precision. A Brier Score of 0.1346 was also obtained, which indicates that it predicts the right behavior 90% of the time.

### *2.4 Multilayer Perceptron in HAR*

Mo et al. [20] stated that the classification of activities based on the CAD-60 Dataset compares convolutionary neural networks and multilayer perceptron efficiency. The CAD-60 Dataset [21] provides RGB-D video sequences of events undertaken by volunteers. Sensor signals are recorded by the Microsoft Kinect sensor. In order to produce highly precise performance results, this research focuses on data preprocessing along with feature extraction. By using CNN for feature extraction and using MLP for the classification of the operation, the model presented incorporates CNN and MLP. It proved highly successful with the model achieving 81.8% accuracy across twelve different types of activities. Catal, Tufekci, Pirmit, and Kocabag [22] compare the performance of a model integrating aspects of decision tree, multilayer perceptron, and logistic regression. Accelerometer data is analyzed to classify activities. Related work [23] performs analysis on the Wireless Sensor Data Mining (WISDM Dataset) which contains information from 36 volunteers performing activities as seen in related work [23]. The proposed model achieved state-of-the-art results while achieving a superior performance when compared to a multilayer perceptron approach in related work. Results prove that integrating an ensemble of a classifier yields outstanding results in the activity recognition domain.

Talukdar and Mehta [8] built a multilayer perceptron network to classify physical human activities through the automated analysis of video data. The volunteer performed six activities 25 different times wearing a variation of different clothes each time. The activities performed were; walking, jogging, running, boxing, hand waving, and hand clapping. Talukdar et al. [8] present an MLP network that trains the data through a recurrent neural network that led to a vast reduction in learning time for the features and labels. The model achieved an overall accuracy of 92%. A comparison of the classification efficiency of XGBoost, MLP, CNN, and LSTM is one feature that is absent from the previously mentioned linked work on the machine

and deep learning models. To determine which network best fits the MHEALTH and WISDM dataset, our goal of this article is to conduct an investigation and compare these six different machine and deep learning algorithms with each other. This project revolves around the topic of using deep learning to benefit the healthcare industry.

Remote patient management (RPM) is one feature that could benefit from deep learning. Sufficient real-time surveillance of the actions of distant individuals will provide tremendous benefits in medical settings. Through reviewing, data sent to them by RPM technology, physicians, nurses, and clinicians may establish good relationships with and strengthen the experience of their patients. The data sent to them via RPM, as seen in [2], will establish a customized care plan and participate in shared decision making to encourage better results. Wearable devices can feed data to a clinician in real time by producing this data, leading to a substantial reduction in continuous patient surveillance. This device may be useful for the elderly, those who are vulnerable to heart problems (or severe medical conditions) and those who suffer from chronic illness. According to [4], the most frequent cause of readmission for patients in the USA is chronic heart failure (CHF). It is calculated that up to 84% of readmissions were considered preventable over a seven-day duration, while 76% of readmissions over a 30-day period were still considered preventable [4].

## 3 Proposed System

See Fig. 1.

### *3.1 Convolutional Neural Network*

In a neural network, neurons learn from each other as they are fully connected. Neurons in convolutional neural networks connect to a fraction of the neurons that are in the previous layer. This layer is the receptive field as seen in related work [24]. Neurons in convolutional neural networks have three dimensions. These dimensions are width, height, and depth.

#### 3.1.1 Architecture

CNNs have a unique architecture. It contains many sequential layers such as the convolutional layer, pooling layer, rectified linear unit layer, normalization layer, and fully connected layer.

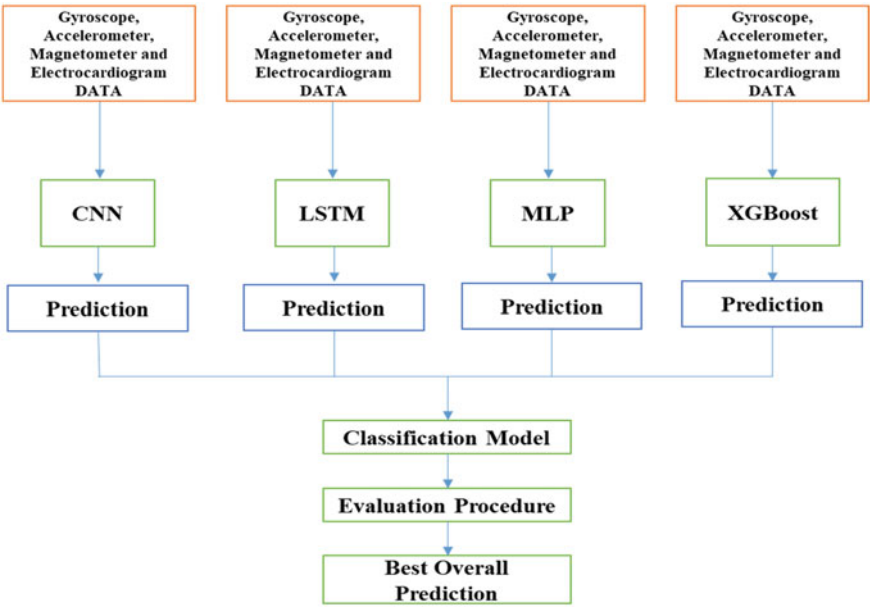



**Fig. 1** Data flow diagram for the proposed model

### 3.1.2 Convolutional Layer

The convolutional layer is the focal point of a CNN. The convolutional layer's main objective is to extract high-level features about the data.

### 3.1.3 MaxPooling Layer

Pooling layers allow for the reduction in a number of parameters in the neural network. It reduces the number of descriptive parameters used to explain the structure of the neural network. It essentially avoids overfitting as it reduces the spatial size of the network. Training a neural network takes a great amount of time. Pooling ensures the number of computations needed to train the network is minimized. It ensures the classification task runs smoothly.

### 3.1.4 ReLU Layer

Relations in data are often nonlinear. The ReLU layer ensures there is an increase in nonlinearity. It applies the following element-wise non-saturating activation function to ensure that a neural network can build the nonlinear relation between data points. If there were no ReLU layer, a neural network would not be able to classify nonlinear

data points. When compared to tanh and sigmoid, the ReLU layer prevails in terms of speed, accuracy, and precision. The width, height, and depth of the neural network, also known as the spatial size, are left unchanged.

### 3.1.5 Dropout

Overfitting is a common problem neural networks face when training data. The dropout regularization technique successfully prevents overfitting. Fully connected layers in a neural network have many variations. Dropout identifies the nodes in a specific layer and removes them. A definitive probability, p, is then applied to the layer. The training process removes nodes linked with the removed layer. As seen in [25], after training, these nodes are placed back into the neural network and assigned their original value (weight). This, in turn, boosts the performance of the neural network. The validation training set benefits the most from this during the deployment of the model.

### 3.1.6 Optimizer: Adam

Adam is a gradient-based optimizer. It is straightforward, simple to implement, and is computationally inexpensive. It is suited to solving classification problems related to human activity recognition. The data involved in HAR is normally relatively large, leading to Adam to be a perfect fit. The hyperparameters require little or no tuning, which is why Adam is the most common optimizer in convolutional neural networks.

### 3.1.7 Softmax Activation Function

The softmax activation function is usually set in the output layer and loss layer. This is usually the final layer in the neural network before the output layer presents the result. The following equation is a detailed representation of the softmax activation function. The layers described above make up the full architecture of a convolutional neural network. The equation below incorporates all the layers and functions mentioned above to represent the typical architecture of a CNN (Fig. 2).

Feature extraction, encoding the labels to one-hot type along with separating the training and testing data, is done before the model is initialized.

CNN activity recognition overview:

(a) One input layer containing 23 features.
(b) Two separable convolution 1D layers with max pooling.
(c) Three hidden layers that were big enough to train the data well.
(d) Two dropout layers that yielded positive results.
(e) Quite good performance, moderately slow.
(f) One output layer of 12 results (labels).

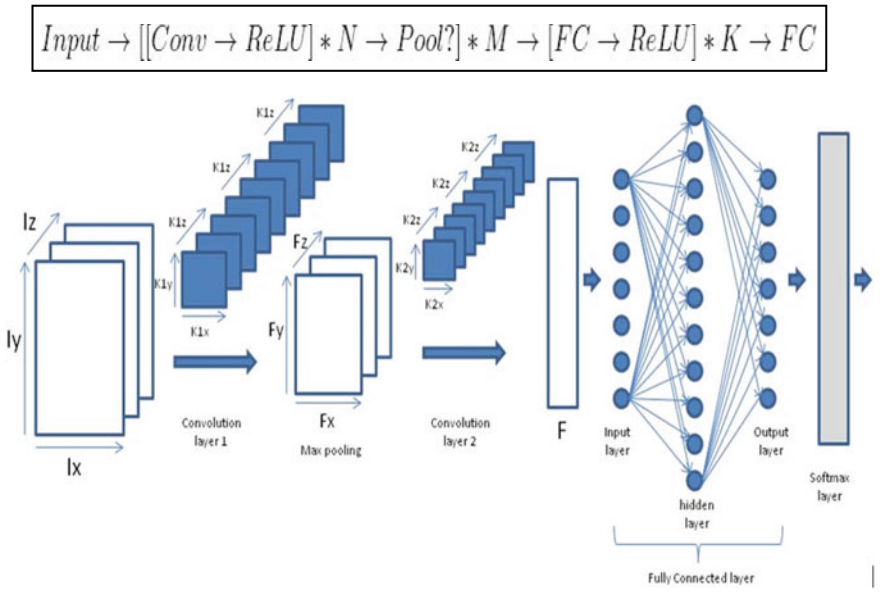



**Fig. 2** Typical architecture of a CNN [26]

To classify (predict) the class variable, which is the motion that each subject executes, the neural network found in the CNN model utilizes the data values given for each of the 23 signals reported. This section gives an overview of training the model and hyperparameter setting.

(a) In the training process, the 'fit()' function is used to train the CNN model.
(b) 'X_train' represents the training data.
(c) 'y_train' refers to the target data.
(d) 'X_test, y_test' represent the validation data.
(e) The model is trained on a total of 245,584 parameters.
(f) While training the model:
(g) The Learning rate is set as 0.0005.
(h) Batch size is set as 32.
(i) The training process is run for 20 epochs.

### 3.1.8 Algorithm Speed

CNN performed excellently on this classification problem. It processed the data relatively fast. Taking into account the speed of the other deep learning algorithms and considering the performance is achieved, the time the model took to train the data was 242 min 18 s.

## *3.2 Long Short-Term Memory*

LSTM activity recognition overview:

(a) One input layer containing 23 features.
(b) Two LSTM layers.
(c) Two dropout layers.
(d) Three hidden layers and one output layer.
(e) One output layer of 12 results (labels).

This section gives an overview of training the model and hyperparameter setting.

(a) In the training process, the 'fit()' function is used to train the CNN model.
(b) 'X_train' represents the training data.
(c) 'y_train' refers to the target data.
(d) 'X_test, y_test' represent the validation data.
(e) The model is trained on 175,373 parameters. While training the model:
(f) The Learning rate is set as 0.0005.
(g) Batch size is set as 32.
(h) Training process is run for 20 epochs.

### 3.2.1 Algorithm Speed

LSTM performed very poorly on the classification problem. It processed the data very slowly. Upon evaluating all six algorithms, LSTM is the poorest performing algorithm and processes data ten times slower than the other algorithms. The total amount of time the model took to process the data was 10 h. At first, the LSTM model was set to run for 20 epochs. The hard drive used to conduct each experiment was not strong enough to process data for such a significant amount of time. Considering the level of performance and speed of processing, CNN was the poorest algorithm applied to the MHEALTH dataset.

## *3.3 Extreme Gradient Boosting*

Gradient boosting machines are associated with a distinctive type of machine learning branch called ensemble learning. The objective of ensemble learning is to train and predict a variety of models at the same time, while each model aims to outperform each others with respect to their output. Consider for example, the route from Paris to Berlin. There are many alternative travel options. As you proceed to take each route, you begin to learn which route is faster and more efficient, leading to the 'superior' route. Taking time to learn, each model has led to the conclusion that X is the superior route. Ensemble learning simply implies this strategy.

XGBoost implements the boosting. Boosting aims to convert weak learners to strong learners. During boosting, iterations lead to the weights of weak learners to

adjust accordingly. Bias reduces allowing for an increase in performance. Accuracy, precision, and recall benefit from the implementation of boosting greatly, as well as a range of evaluation techniques. Extreme gradient boosting (XGBoost) is the best performing boosting algorithm. XGBoost is a decision-tree-based algorithm that utilizes the use of gradient boosting and ensemble learning. XGBoost performs so well on data due to its ability to transform weak learners into strong learners. It utilizes boosting within the gradient descent architecture. It allows the framework to develop dramatically with its quick and easy to learn optimization techniques and parameter enhancements.

Feature extraction and splitting of the training and testing data are performed before the model is initialized. XGBoost activity recognition overview:

(a) Multiclass classification using the softmax activation function.
(b) Evaluation: multiclass classification error rate.
(c) Learning rate is set to 0.05.
(d) Trained on 161,959 parameters.

The parameters regarding the XGBoost model are discussed.

(a) The model is trained using the parameter list and the training data.
(b) The model is trained for 10 rounds.
(c) Early stopping.
(d) The learning rate of the model is 0.05 while the number of estimators is set to 1000.
(e) Early stopping is used in the validation set to identify the appropriate amount of boosting rounds. This is usually the optimal number of boosting rounds needed. It is set to stop within 5 rounds. It will train until 'validation_0-merror' has not improved in 5 rounds.
(f) The XGBoost model will train until the 'validation_0-merror' has not improved anymore. The 'validation_0-merror' must continue to decrease in order for the 'early_stopping_rounds' to continue to train the XGBoost model.

### 3.3.1 Algorithm Speed

XGBoost performed very well on this classification problem. It processed the data relatively fast. Taking into account the speed of the other deep learning algorithms and considering the performance it achieved, XGBoost was the second-best performing algorithm.

## *3.4 Multilayer Perceptron*

Feature extraction, encoding the labels to one-hot form along with splitting the training and testing data, is performed before the model is initialized. MLP activity recognition overview:

(a) One input layer containing 23 features.
(b) Four hidden layers that were big enough to train the data well.
(c) Two dropout layers that benefited the results greatly. The top-class algorithm is based on the concept of gradient descent.
(d) One output layer of 12 results (labels).

This section gives an overview of training the model and hyperparameter setting.

(a) In the training process, the 'fit()' function is used to train the CNN model.
(b) 'X_train' represents the training data.
(c) 'y_train' refers to the target data.
(d) 'X_test, y_test' represent the validation data.
(e) The model is trained on 706,317 parameters. While training the model:
(f) The learning rate is set as 0.0005.
(g) Batch size is set as 32.
(h) Training process is run for 20 epochs.

#### 3.4.1 Algorithm Speed

MLP performed excellently on this classification problem. It processed the data relatively fast. Taking into account the speed of the other deep learning algorithms and considering the performance it achieved, the time the model took to train the data was 86 min 40 s.

### *3.5 MHEALTH Dataset*

The MHEALTH dataset consists of body motion and vital signs recordings. Ten volunteers conducted the experiment, each with different characteristics. The subjects' task is to perform 12 different types of activities. The accelerometer, gyroscope, and magnetometer placed on the subjects' body measure acceleration, rate of turn, and magnetic field orientation. These sensors measure the range of motion experienced by each body part. The electrocardiogram sensor positioned on the chest also provides 2-lead ECG measurements. ECG can assist in the basic heart monitoring, checking for various arrhythmias, or looking at the effects of exercise on the ECG.

## 4 Result and Performance Analysis

Accuracy, precision, recall, F1-score of each DL architecture. Figure 1 outlines the accuracy of the MLP, XGBoost, and CNN machine learning models on the

MHEALTH dataset. The figures are extracted from the model's related confusion matrix. MLP attains the highest values, followed by MLP, then CNN (Table 1).

The four classification models presented in this research perform well when compared to existing state-of-the-art baselines. MLP and XGBoost achieve excellent performance measures, challenging many research papers with improved accuracy, precision, recall, and F1-score. The XGBoost model is the best performing model in terms of overall performance and is highly suited to mobile health data.

Figure 3 outlines the accuracy of the MLP, XGBoost, and CNN machine learning models on the MHEALTH dataset. The figures are extracted from the model's related confusion matrix. MLP attains the highest values, followed by MLP, then CNN (Table 2).

Figure 4 examines the performance of each deep learning model on the classification task, the figures are taken from each approaches confusion matrix output. MLP performs the best in comparison to the other models. The CNN and LSTM models achieved an average performance of 66.2% and 48.9%, respectively. The hybrid and

**Table 1** Performance metrics evaluation

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score |
|---|---|---|---|---|
| MLP | 90.55 | 91.66 | 90.55 | 90.7 |
| CNN | 83.91 | 83.47 | 83.91 | 82.98 |
| LSTM | 78.09 | 74.86 | 78.09 | 75.6 |
| XGB | 89.97 | 90.09 | 89.97 | 89.78 |

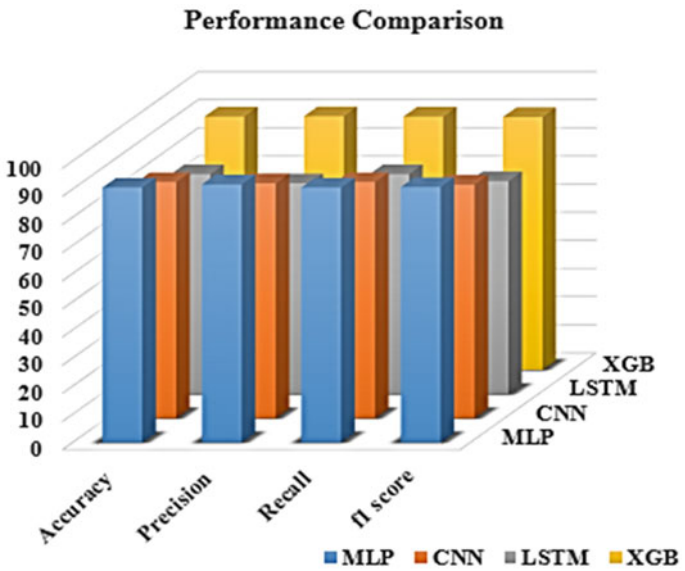



**Fig. 3** Performance comparison

**Table 2** Performance metrics evaluation

| Activity | CNN | LSTM | XGBoost | MLP |
|---|---|---|---|---|
| Standing still | 95 | 63 | 94 | 97 |
| Sitting and relaxing | 100 | 97 | 96 | 99 |
| Lying down | 100 | 100 | 98 | 100 |
| Walking | 45 | 0 | 77 | 96 |
| Climbing stairs | 38 | 0 | 48 | 92 |
| Waist bends forward | 81 | 53 | 76 | 96 |
| Frontal elevation of arms | 86 | 70 | 84 | 96 |
| Knees bending | 68 | 38 | 67 | 93 |
| Cycling | 33 | 42 | 92 | 97 |
| Jogging | 75 | 47 | 89 | 98 |
| Running | 61 | 68 | 92 | 91 |
| Jump front and back | 12 | 9 | 62 | 67 |
| Average | 66.2 | 48.9 | 81.25 | 93.5 |

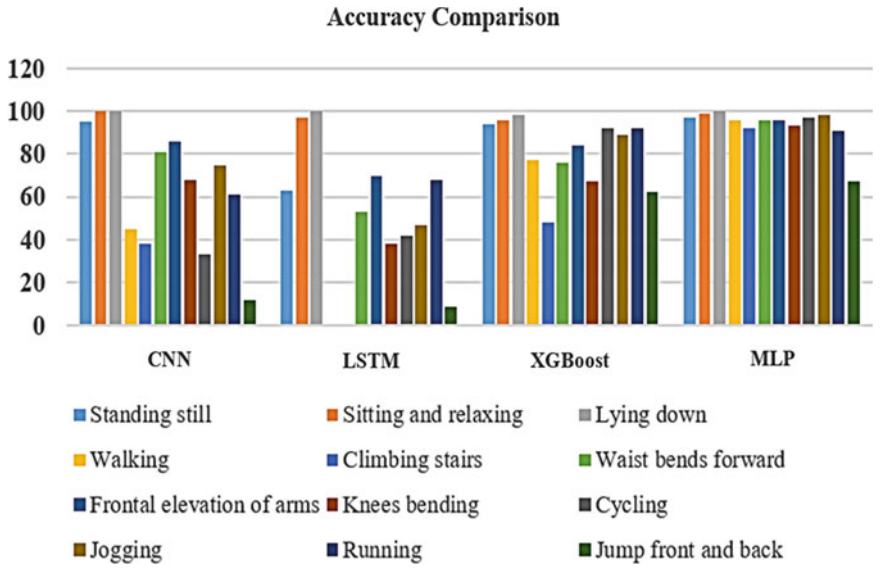



**Fig. 4** Accuracy comparison

XGBoost models achieved an average performance of 70.6% and 81.25%, respectively. For each subject, the MLP model attained the best performance obtaining an average performance of 93.5%. The figure identifies the superiority of the MLP model over others by showing a significantly higher performance.

### *4.1 Accuracy and Loss Results*

Fine-tuning each model is to classify the data accurately, which is significant to maximizing accuracy while minimizing loss. MLP, XGBoost, CNN, and LSTM are fine-tuned with hyperparameters, respectively. The loss and accuracy of each model are visualized throughout this section. After fine-tuning each model and using a total of 20 epochs, MLP prevailed as the network with the highest accuracy (91%) and minimal loss. The average accuracy for each model is relatively high. MLP and XGBoost performed better and more consistently than CNN, LSTM. MLP and XGBoost achieved accuracies of 91% and 89%, respectively. MLP and XGBoost were able to converge far more easily as outlined in Fig. 3 which depicts the multilayer perceptron model achieving 91% accuracy.

The following two plots depict the convolutional neural network model achieving 84% accuracy. CNN performed moderately well in terms of accuracy (84%) and loss. MLP and XGBoost increased their accuracy along with diminishing their loss as the number of training iterations increased. Overall, MLP outperformed the other networks with the highest accuracy and lowest loss. The LSTM model misclassified too many instances, which lead it to become the poorest performing model with the least accuracy (78%) and highest loss.

LSTM networks performed poorly with 20 training iterations as depicted in Figs. 5. They both achieved 78% and 84% accuracy, respectively. The following two plots depict the LSTM model achieving 78% accuracy (Figs. 6 and 7).

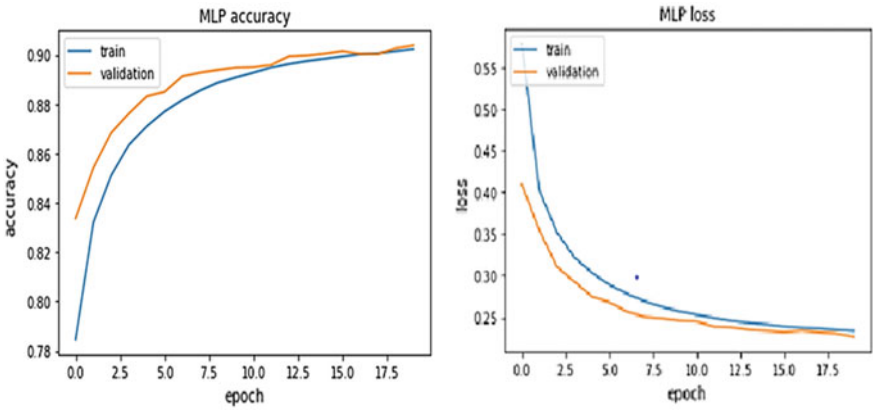



**Fig. 5** MLP accuracy and loss

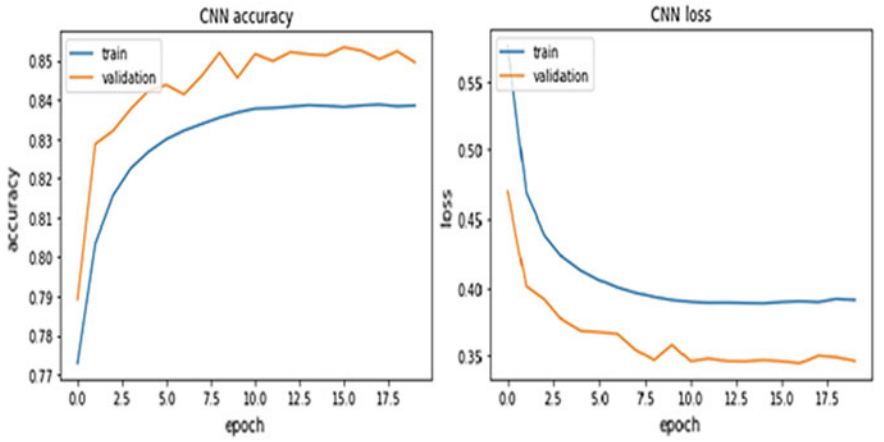



**Fig. 6** CNN accuracy and loss

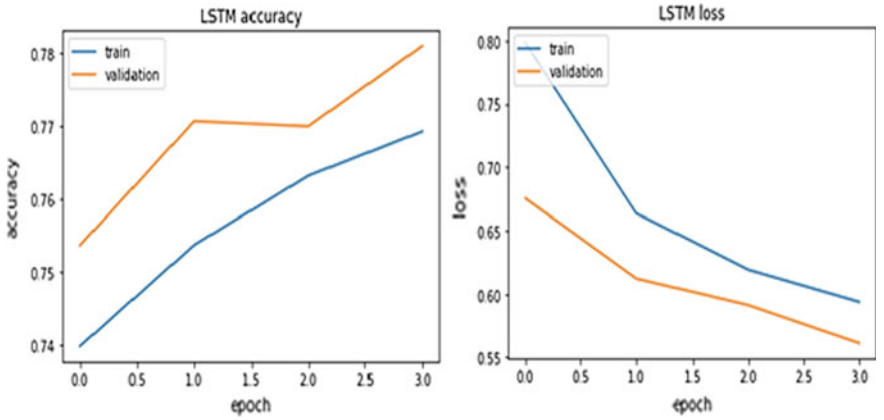



**Fig. 7** LSTM accuracy and loss

## 5 Conclusion and Future Enhancement

Each model performs human activity recognition from wearable sensors such as gyroscopes, accelerometers, magnetometers, and electrocardiograms. To the author's knowledge, for the MHEALTH dataset, XGBoost has not been performed to classify the activities in question. MLP prevailed as the best performing model achieving accuracy, precision, recall, and F1-score of 90.53%, 91.71%, 90.53%, and 90.76%, respectively. XGBoost was the next best performing model that achieves accuracy, precision, recall, and F1-score of 89.98%, 90.14%, 89.98%, and 89.78%, respectively. While MLP outperformed XGBoost in terms of precision, accuracy, recall, and F1-score, 471 instances were misclassified by MLP, while XGBoost misclassified just

281.1341, 2533, and 2742 instances were misclassified, respectively, by CNN, and LSTM. XGBoost is the highest performing model in terms of total precision, consistency, recall, F1-score, and number of correctly categorized instances. This outlines the established area of appropriateness for the XGBoost system, which was never documented using wearable sensors on the MHEALTH dataset. This section presents an account of future work in human activity recognition using deep learning. Some of the ways in which human activity recognition models using deep learning will benefit healthcare in remote patient monitoring: clinician decision support, ambient assisted living / aiding the elderly, drug discovery, developing regions whose healthcare services are limited, app with patient data, diagnostic abilities, reduce need for electronic health records, creating more precise analytics for diagnosis, clinical decision making, risk scoring, and early alerting. Common challenges on human activity recognition presented by this research are unmeasurable uncertainty factors, activity similarity, and the null class problem.

## References


1. Heart.org (2019) [Online]. Available: https://www.heart.org/-/media/files/about-us/policy-research/policy-positions/clinical-care/remote-patient-monitoring-guidance-2019.pdf
2. Khusainov R, Azzi D, Achumba IE, Bersch SD (2013) Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations. Sensors
3. Roobini S, Fenila Naomi J (2019) Smartphone sensor based human activity recognition using deep learning models. Int J Recent Technol Eng (IJRTE) 8(1)
4. Wang J, Chen Y, Hao S, Peng X, Hu L (2019) Deep learning for sensor-based activity recognition: a survey. Pattern Recog Lett 119:3–11
5. Slim SO, Atia A, Elfatta MM, Mostafa MSM (2019) A survey on human activity recognition based on acceleration data. Int J Adv Comput Sci Appl 10:84–98
6. Sousa W, Souto E, Rodrigres J, Sadarc P, Jalali R, El-Khatib K (2017) A comparative analysis of the impact of features on human activity recognition with smartphone sensors. In: Proceedings of the 23rd Brazillian symposium on multimedia and the Web, Gramado, Brazil, 17–20 Oct 2017; pp 397–404
7. Foerster F, Smeja M, Fahrenberg J (1999) Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Comput Hum Behav 15(5):571–583
8. Bashar A (2019) Survey on evolving deep learning neural network architectures. J Artif Intell 1(2):73–82
9. Prabakeran S (2018) In-depth survey to perceiving the effect of kidney dialysis parameters using clustering framework. J Comput Theor Nanosci 15(6–7):2233–2237
10. Indumathi V (2018) Utilizing data mining classification technique to predict kidney diseases. J Comput Theor Nanosci 15(6–7):2193–2196
11. Jiang W, Yin Z (2015) Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM international conference, pp 1307–1310
12. Banos O, Garcia R, Saez A (2019) UCI machine learning repository: MHEALTH Dataset Data Set, Archive.ics.uci.edu
13. Zhang M, Sawchuk AA (2012) Human activities dataset. Sipi.usc.edu
14. Shoaib M, Bosch S, Durmaz Incel O, Scholten H (2014) Fusion of smartphone motion sensors for physical activity recognition. Sensors
15. Hammerla NY, Halloran S, Plotz T (2016) Deep, convolutional, and recurrent models for human activity recognition using wearables. IJCAI

16. Kim Y, Moon T (2015) Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks
17. Javier Ordóñez F (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors 16:115 [69]
18. Zhang W, Zhao X, Li Z (2019) A comprehensive study of smartphone-based indoor activity recognition via Xgboost. IEEE Access 7:80027–80042
19. Nguyen T, Fernandez D, Nguyen Q, Bagheri E (2017) Location-aware human activity recognition. Adv Data Min Appl 821–835
20. Mo L, Li F, Zhu Y, Huang A (2016) Human physical activity recognition based on computer vision with deep learning model. In: 2016 IEEE international instrumentation and measurement technology conference proceedings
21. Cornell Activity Datasets: CAD-60 & CAD-120 | re3data.org, Re3data.org, 2019
22. Catal C, Tufekci S, Pirmit E, Kocabag G (2015) On the use of ensemble of classifiers for accelerometer-based activity recognition. Appl Soft Comput
23. Talukdar J, Mehta B (2019) Human action recognition system using good features and multilayer perceptron network
24. Luo W, Li Y, Urtasun R, Zemel R (2019) Understanding the effective receptive field in deep convolutional neural networks
25. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
26. Mody M, Mathew M, Jagannathan S, Redfern A, Jones J, Lorenzen T (2019) CNN inference: VLSI architecture for convolution layer for 1.2 TOPS. Ieeexplore.ieee.org

# FETE: Feedback-Enabled Throughput Evaluation for MIMO Emulated Over 5G Networks


**B. Praveenkumar, S. Naik, S. Suganya, I. Balaji, A. Amrutha, Jayanth Khot, and Sumit Maheshwari**

**Abstract** Mobile networks are playing a tremendous role in our day-to-day activities. In the currently evolving networks such as 5G, satisfying quality of service (QoS) is remaining as a challenging problem due to the dense network deployment. Moreover, multiple technologies such as LTE, Wi-Fi, and 5G contending and cooperating make the resource allocation a complex problem. This paper attempts to optimize the radio resource allocation in heterogeneous wireless networks for a particular geographical region by finding the throughput by maintaining the QoS along with a combination of network parameters. Specifically, the proposed research work uses different parameters such as RSSI, RSRP, and RSRQ for calculating the throughput of user equipment in a specified area. The feedback-enabled method (FETE) is then compared with and evaluated for the MIMO system, where it is observed that an overall throughput gain can be obtained by using right optimization technique for different parameters.

**Keywords** MIMO · LTE · 5G · Network evaluation · Throughput · Scheduling

## 1 Introduction

The tremendous growth of 5G will be due to its better performance in terms of capability, capacity, data rate, and latency when compared with other co-existent technologies such as 3G, LTE and LTE-advanced, is incomparable. The rapidly evolving 5G networks have the risk of overcrowding the frequency range of wireless spectrum as a multitude of devices that attempt to connect to a single frequency channel. Furthermore, the overly dense deployment of 5G will present complex challenges, when accompanied by the heterogeneity, scale, and diverse quality of service (QoS) requirements.

B. Praveenkumar (✉) · S. Naik · S. Suganya · I. Balaji · A. Amrutha · J. Khot
Electronics and Communication Department, CMR Institute of Technology, Bengaluru, India

S. Maheshwari
Electrical and Computer Engineering, WINLAB, Rutgers University, North Brunswick, USA



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_71

Existing techniques of throughput optimization depend upon maximizing the usage of available network resources. However, due to lack of optimizations based on a geographical region while using MIMO leave some resources non-utilized or under-utilized specifically when paired with the user mobility. Therefore, there is a need to use better techniques which can support the increase in the overall throughput received by the user equipment (UEs) which are mobile. Several schemes are proposed in the literature for improving throughput using different parameters such as channel quality indicator (CQI), round-robin (RR) scheduling for UEs resource blocks (RBs), first in first out (FIFO) based on UE connections, and blind equal throughput (BET) to achieve fairness [1]. Despite these approaches, the QoS is always constrained due to inherently low resources available to fulfill the usage requirements specifically when targeting high throughput and low-latency applications. The resource scarcity in part is addressed using relatively newer efficient solutions that optimize the use of resources based on machine learning and other estimation techniques. Depending on the QoS, these techniques can be classified as user-based or network-based or both.

The existing techniques among other factors also lack an integrated approach or the systems view when dealing with the scarce resources. For example, the BET scheduler aims to provide equal throughput to all UEs under an eNB which in turn lacks the fairness or support to a specific application that is resource hungry. The RR scheduling or best CQI again fails to enhance the throughput of the UEs when additional inputs are required to be captured to provide a holistic view on the network.

To optimize upon the available resources and provide an optimal output for the throughput using the inputs from UEs in a particular area, a closed-loop feedback-based system is essential. In such a system, the additional inputs can be seamlessly added that play critical role in increasing the overall throughput and supporting low-latency. Supporting these views of inputting additional inputs to improve the throughput, this paper proposes a way to optimize the mobile networks for a multiplicity of parameters for the mobile UEs. Furthermore, 5G studies are incomplete without a discussion on MIMO. Therefore, the system is evaluated for the MIMO and show that additional improvements can be obtained using the same. Figure 1 shows that there is a potential of throughput improvement when considering multiple parameters along with the MIMO capabilities when a feedback-based mechanism is used. This paper proposes FETE, a feedback-enabled throughput evaluation method for the MIMO systems.

The rest of the paper is organized as follows. Section 2 summarizes the existing techniques used previously in the literature. Section 3 illustrates the experiment methodology to evaluate the throughput performance. Section 4 details the simulations carried out to calculate the throughput, with or without MIMO by injecting a number of users. The evaluation results are presented in Sect. 5. Sect. 6 concludes the paper with a view on our future work.

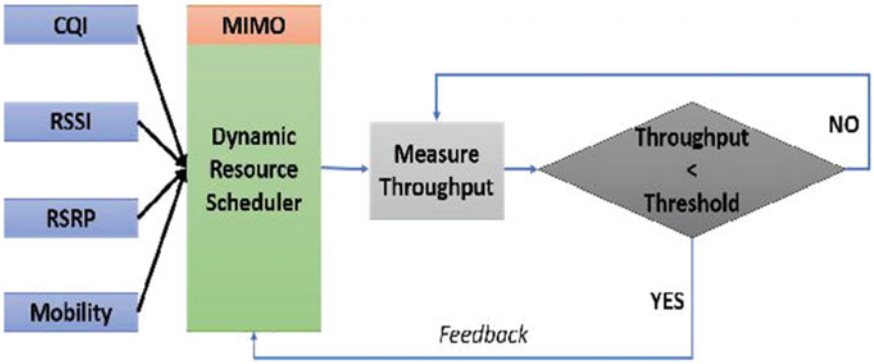



**Fig. 1** Feedback-based throughput improvement in 5G

## 2 Background

Resource scheduling is a vast field. On the one hand, the existing work deals with the multi-parametric approaches to optimize throughput using methods such as estimation [2], prediction [3], resource allocation, machine learning, fuzzy logic [4, 5], and so on. While on the other hand, research has also focused on measurement wherein different algorithms are compared at different settings to find the best possible combination of inputs for a desired output [6, 7].

Currently, the work has evolved when MIMO has taken the precedence over the existing techniques. With the combination of multiple input and output antenna arrays, it is feasible to achieve the higher throughput than the traditional transceivers when the beams are properly aligned. Moreover, using the software-based MIMO enabled using the mm wave [8, 9], it is possible to create a fully software-defined system, thus able to control and coordinate from anywhere.

The software-based automation has provided more meaningful avenues for optimization than ever. The controllable experiments and practical systems can now run under a variety of settings that can be triggered or trimmed in short time. This flexibility encourages us to evaluate the feasibility of having a MIMO system that can be analyzed for different that can be dynamically changed to provide a real-time feedback to the system based upon user mobility and network fluctuations.

This paper evaluates the MIMO system for various parameters. The experiment considers the UE mobility and correlated with the network parameters to calculate the throughput while keeping QoS under considerations. Further,

the system is evaluated with and without MIMO using a real mobility dataset. Our work is distinguished from earlier approaches as it lies at the intersection of evaluation-based enhancement for MIMO that is a key design choice in the 5G networks.

## 3 Methodology

This section provides the method used in this work to calculate the throughput.

### *3.1 Dataset Description*

In this paper, vehicular movement data collected and emulated using ETSI API for 5G using an LTE eNB at the University of Hertfordshire are used. The various scrambling codes are as follows. For uplink, it is 54 with the target SIR of 17.3. The minimum uplink channelization code length is 8. The downlink scrambling code is 1 with the channelization code of 15. The maximum downlink power is 10.1 and minimum is 9.3 dBm. On a pre-set path, the UEs move across the eNB, wherein the RSSI and throughput are collected for each time tick. The GPS coordinates are also logged for each location at the time tick. Thus, for various UEs, the real dataset is obtained along with their location with respect to time.

### *3.2 Exploration and Extrapolation*

The dataset lacks the measurements at some of the GPS locations for some UEs. Therefore, before using, the data are cleaned by removing any anamoly, and for some points, the throughput is calculated at a given location by considering the available parameters like RSSI, RSRP, and RSRQ whichever were available.

### *3.3 Analytical Model*

The parameters obtained from the dataset are plotted as follows.

RSRP: Reference signal received power (RSRP) is defined as the linear average over the power contributions (in [W]) of the resource elements that carry cell-specific reference signals within the considered measurement frequency bandwidth. UE measures the power of multiple resource elements used to transfer the reference signal but then takes an average of them rather than summing them. Figure 2 shows the received RSRP (in dB) for a sample mobile UE.

RSRQ: Reference signal received quality (RSRQ) is a C/I type of measurement, and it indicates the quality of the received reference signal. The RSRQ measurement provides additional information when RSRP is not sufficient to make a reliable handover or cell re-selection decision. Figure 3 shows the RSRQ (in dB) for a sample mobile UE connected with a base station at different time ticks.

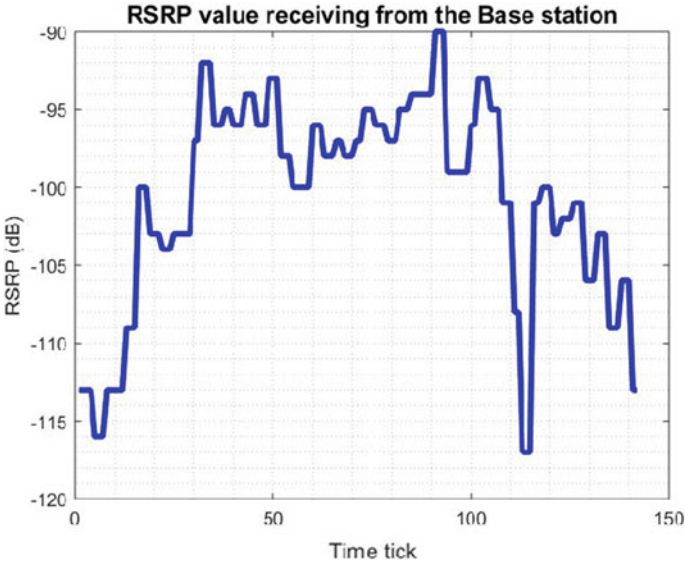



**Fig. 2** RSRP values received at a UE from the base station

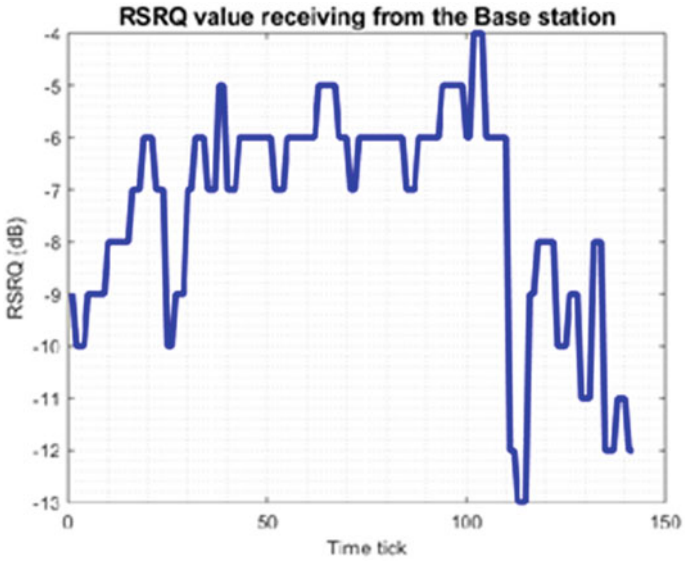



**Fig. 3** RSRQ values received at a UE from the base station

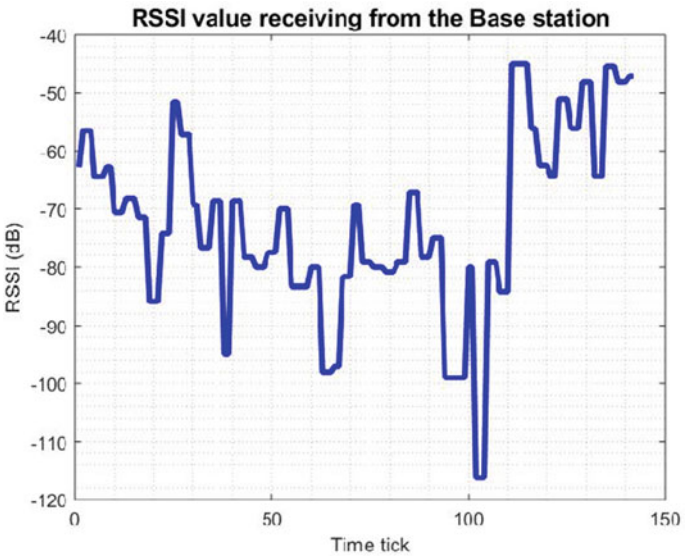



**Fig. 4** RSSI values for a sample UE from the base station

RSSI: The carrier receive strength signal indicator (RSSI) measures the average total received power observed only in the OFDM symbols containing reference symbols for antenna port 0 (i.e., OFDM symbol 0 and 4 in a slot) in the measurement bandwidth over N resource blocks. The total received power of the carrier RSSI includes the power from co-channel serving and non-serving cells, adjacent channel interference, thermal noise, etc.

The total RSSI is measured over 12 subcarriers including the signals from serving cell, and the traffic in the serving cell. The relationship between RSSI, RSRP, and RSRQ is given in Eq. 1 where N are the number of resource blocks.

$$\text{RSSI} = N(\text{RSRP} = \text{RSRQ}) \tag{1}$$

Figure 4 shows the calculated RSSI values using Eq. (1)

However, the relative humidity is proportional to the RSSI: When relative humidity increases, RSSI level increases as well. The impact of temperature and relative humidity variation on the RSSI is stronger when the distance between the sender and the receiver is large.

## *3.4 Signal Continuity*

After obtaining RSSI values at discrete locations, 50 users are injected at the random coordinates in the map area. The RSSI values of these users are correlated with

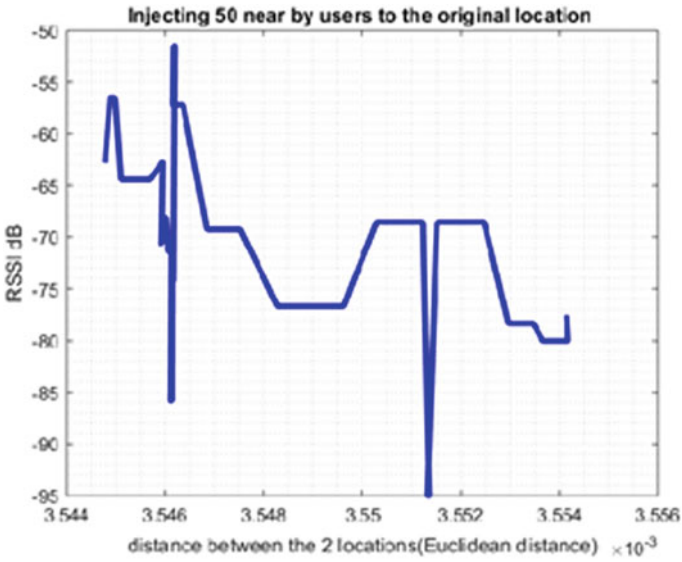



**Fig. 5** Euclidean distance versus RSSI for different users

the RSSI dataset by matching them with the closest geographical location and incorporating error corrections.

It can be easily plotted with RSSI value along the *Y*-axis and Euclidean distance between original location and injected user location along the *X*-axis. Figure 5 plots the Euclidean distance between any two users in the system and their respective RSSI value. This provides us a capability of obtaining RSSI at any given location in the map.

## 4 Feedback-Enabled Throughput Evaluatıon

The feedback-enabled throughput evaluation (FETE) develops a mechanism through which the continuous values of throughput can be obtained in a closed MIMO system.

### *4.1 System Informatıon*

Using the RSSI values for 50 users, the downlink data rate is calculated using transitional CQI value.

CQI: The channel quality indicator (CQI) contains information sent from a UE to the eNB/gNB to indicate a suitable downlink transmission data rate, i.e., a modulation and coding scheme (MCS) value. CQI is a 4-bit integer and is based on the observed

**Table 1** Modulation scheme based on the CQI value

| RSSI | CQI | Modulation scheme |
|---|---|---|
| Excellent | 10–15 | 64 QAM |
| Good | 7–9 | 16 QAM |
| Fair | 1–6 | QPSK |
| Poor | 0 | Out of range |

**Table 2** Bandwidth based on the number of resource blocks

| Number of resource blocks | Channel bandwidth (MHz) |
|---|---|
| 6 | 1.4 |
| 15 | 3 |
| 25 | 5 |
| 50 | 10 |
| 75 | 15 |
| 100 | 20 |

signal-to-interference-plus-noise ratio (SINR) at the UE. The CQI estimation process takes into account the UE capability such as the number of antennas and the type of receiver used for detection.

Using the RSSI values obtained using the RSRQ and RSRP, the CQI values are determined using Table 1. The CQI values are granulated using the RSSI values to determine if a signal is good or bad for the UE. Based on the CQI values, the modulation scheme is assigned ranging from 64-QAM, to 16-QAM and QPSK. For a low CQI value, corresponding to the poor RSSI range, the UE is considered as out of range. To calculate the modulation scheme given in Table 1, the channel bandwidth is to be considered as provided in Table 5 (Table 2).

Based on the modulation scheme, a modulation index is assigned by the eNB/gNB, and according to that index value, a respective transport block size (TBS) index value is assigned as given in Table 3.

Finally, for the throughput calculation, the allowed rate of bits in each subframe for the given TBS index and number of physical resource blocks (PRBs) is to be determined. The bits/subframe is allocated by the eNB/gNB based on the traffic conditions, UE category (priority, services, etc.), and RF conditions. The standard bits/subframe allocation as per 3GPP for different RBs and TBS index value is given in Table 3 for 50 and 100 RBs, respectively.

## *4.2 Throughput Calculation*

Using the MCS index, TBS index, CQI value, and PRBs, the throughput is calculated as follows.

**Table 3** Modulation index and TBS index

| Modulation Index | Modulation | TBS index |
|---|---|---|
| 0 | QPSK | 0 |
| 1 | | 1 |
| 2 | | 2 |
| 3 | | 3 |
| 4 | | 4 |
| 5 | | 5 |
| 6 | | 6 |
| 7 | | 7 |
| 8 | | 8 |
| 9 | | 9 |
| 10 | 16 QAM | 9 |
| 11 | | 10 |
| 12 | | 11 |
| 13 | | 12 |
| 14 | | 13 |
| 15 | | 14 |
| 16 | | 15 |
| 17 | 64 QAM | 15 |
| 18 | | 16 |
| 19 | | 17 |
| 20 | | 18 |
| 21 | | 19 |
| 22 | | 20 |
| 23 | | 21 |
| 24 | | 22 |
| 25 | | 23 |
| 26 | | 24 |
| 27 | | 25 |
| 28 | | 26 |
| 29 | QPSK | Reserved |
| 30 | 16 QAM | Reserve d |
| 31 | 64 QAM | Reserved |

For example, when the RSSI value is excellent, the 64-QAM modulation can be chosen as per Table 1. For the 64-QAM, a granulated MCS index value is allocated by the eNB/gNB along with the PRBs. Consider, for example, that the MCS index is 28 in which case the respective TBS index value of 26 will be chosen as given in Table 3. For a system with 50 PRBs, for the chosen 26 TBS index value, 32,856 bits/subframe (1 ms) is allocated to the UE. In turn, a throughput of 32,856 1000 = 32,856,000 bits/sec, i.e., 32.85 Mbps can be obtained. In this particular example, for a 10 MHz bandwidth, it obtains around 32Mbps throughput, when assuming that all 50 resource PRBs are allocated to a single UE.

### *4.3 Improved Throughput Using MIMO*

The system efficiency can be enhanced or it can increase the throughput with the help of MIMO technology.

MIMO: Multiple input multiple output (MIMO) allows multiple inputs and output antenna to be used simultaneously, thus theoretically improving the capacity manifolds. Consider that 101,101 data is transmitted through a channel with deep fades. Due to the fluctuations in the channel quality, the data stream may get lost or severely corrupted that the receiver cannot recover it. The solution to combat the rapid fluctuation is to add independent fading channels by increasing the number of transmitting or receiver antennas or both. The use of spatial diversity technique, where same information sent or receive across independent channel can help to combat fading, is particularly useful in MIMO. The diversity gain in MIMO is obtained using Eq. (2) where $N_{\text{tx}}$ and $N_{\text{rx}}$ are the number of transmitting and number of receiving antennas, respectively.

$$G_{\text{diversity}} = N_{tx} + N_{rx} \tag{2}$$

In an ideal condition, for $N_{\text{tx}} = 2$ and $N_{\text{rx}} = 2$, diversity gain is 4. This gain can be obtained by adding the independent fading channel that increases the reliability of the transmission link. Therefore, if this MIMO 2 * 2 technology is used for calculating the same throughput for above mentioned UE, the user will receive around 75 Mbps throughput. The evaluation of throughput with and without MIMO is presented in the next section (Tables 4 and 5).

## 5 Evaluatıon and Results

The simulation is carried out for calculating the throughput for, with and without MIMO technology using FETE. The entire procedure to calculate throughput is written as in the JavaScript language using the Visual Studio Code software. The node package manager (NPM) runtime environment is used to run the code. The

**Table 4** Bits/subframe based on number of physical resource block table

| $TBS_i$ndex | $^{N}P\ RB^{50}$ | $N_{P\ RB}100$ |
|---|---|---|
| 0 | 1384 | 2792 |
| 1 | 1800 | 3624 |
| 2 | 2216 | 4584 |
| 3 | 2856 | 5736 |
| 4 | 3624 | 7224 |
| 5 | 4392 | 8760 |
| 6 | 5160 | 10,296 |
| 7 | 6200 | 12,216 |
| 8 | 6968 | 14,114 |
| 9 | 7992 | 15,840 |
| 10 | 8760 | 17,568 |
| 11 | 9912 | 19,848 |
| 12 | 11,448 | 22,920 |
| 13 | 12,960 | 25,456 |
| 14 | 14,112 | 28,446 |
| 15 | 15,264 | 30,576 |
| 16 | 16,416 | 32,856 |
| 17 | 18,336 | 36,696 |
| 18 | 19,848 | 39,232 |
| 19 | 21,384 | 43,816 |
| 20 | 22,920 | 46,888 |
| 21 | 25,456 | 51,024 |
| 22 | 27,376 | 55,056 |
| 23 | 28,336 | 57,336 |
| 24 | 30,576 | 61,664 |
| 25 | 31,706 | 62,776 |
| 26 | 32,856 | 75,376 |

plotting is done using Highchart and localhost:8000 is used to show the simulated output in a browser.

By injecting 50 users into the map area, the throughput with and without MIMO is calculated. The simulation code is flexible and incorporates multiple parameters in real time. Figure 6 shows the simulation output for a single run for different number of users. It can be observed that the throughput is higher when using MIMO.

Table 5 presents the average throughput calculated for both, with and without MIMO cases for the first nine locations in Fig. 6. The throughput using MIMO is twice of that without the MIMO.

As the MCS index value is randomly chosen, different results are obtained for various simulation runs. Therefore, Fig. 7 shows the average over 15 simulation runs

**Table 5** Average throughput calculation

| Throughput without MIMO | Throughput with MIMO |
|---|---|
| 31,704,000 | 63,408,000 |
| 28,336,000 | 56,672,000 |
| 25,456,000 | 50,912,000 |
| 21,384,000 | 42,768,000 |
| 30,576,000 | 61,152,000 |
| 15,264,000 | 30,528,000 |
| 16,416,000 | 32,832,000 |
| 27,376,000 | 54,752,000 |
| 21,384,000 | 42,768,000 |
| Avg = 21,383,111 (21 Mbps) | Avg = 42,766,222 (42.7 Mbps) |

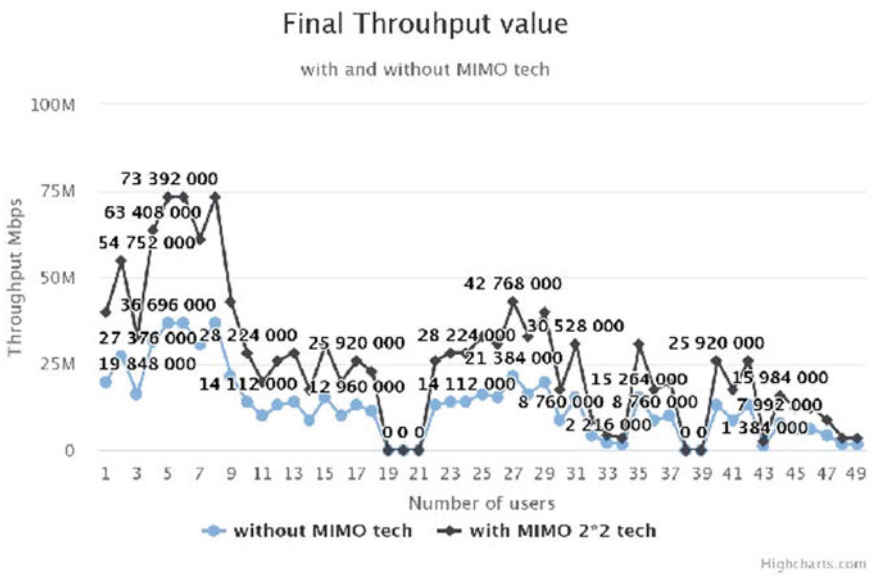



**Fig. 6** Throughput for MIMO and without MIMO for different users

for the throughput with and without MIMO. As it can be observed, the overall gain of the throughput is much higher for the MIMO as compared to the one without it.

If SPS does not receive a reply from a particular contact author, within the time-frame given (usually 72 h), then it is presumed that the author has found no errors in the paper. The tight publication schedule of our proceedings series does not allow SPS to send reminders or search for alternative e-mail addresses on the Internet.

In some cases, it is the contact volume editor or the publication chair who checks all of the PDFs. In such cases, the authors are not involved in the checking phase.

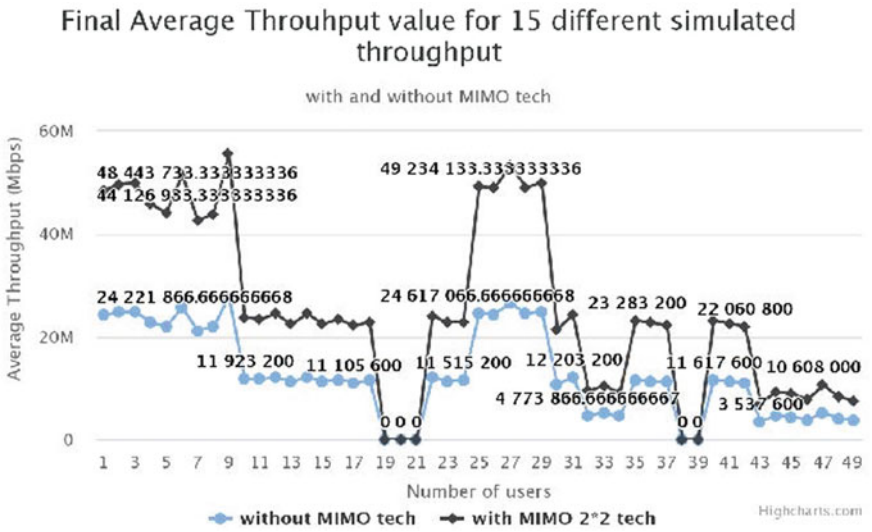



**Fig. 7** Average system throughput comparison

The purpose of the proof is to check for typesetting or conversion errors and the completeness and accuracy of the text, tables, and figures. Substantial changes in content, e.g., new results, corrected values, title, and authorship, are not possible and cannot be processed.

## 6 Conclusion

This paper presented a feedback-enabled throughput evaluation (FETE) method to analyze the 5G network with and without MIMO technique. Using the network-specific parameters such as the MCS index, the TBS index, the CQI values, and the PRBs, the throughput is systematically estimated for a set of mobile UEs that are obtained from a real dataset. With the help of simulation, it has been observed that the overall gain of the throughout is higher for MIMO in comparison to that of the one without MIMO.

## References

1. Suganya S, Maheshwari S, Latha YS, Ramesh C (2016, July) Resource scheduling algorithms for LTE using weights. In: 2016 2nd ınternational conference on applied and theoreticalcomputing and communication technology (iCATccT). IEEE, pp 264–269
2. Chang LF (1991) Throughput estimation of ARQ protocols for a Rayleigh fading channel using fade-and interfade-duration statistics. IEEE Trans Veh Technol 40(1):223–229

3. Maheshwari S, Mahapatra S, Kumar CS, Vasu K (2013) A joint parametric prediction model for wireless internet traffic using Hidden Markov Model. Wirel Netw 19(6):1171–1185
4. Vasu K, Maheshwari S, Mahapatra S, Kumar CS (2011, January) QoS aware fuzzy rule based vertical handoff decision algorithmfor wireless heterogeneous networks. In: 2011 National conference on communications (NCC). IEEE, pp 1–5
5. Tamandani YK, Bokhari MU (2016) SEPFL routing protocol based on fuzzy logic control to extend the lifetime and throughput of the wireless sensor network. Wirel Netw 22(2):647–653
6. Ye W, Heidemann J, Estrin D (2004) Medium access control with coordinated adaptive sleeping for wireless sensor networks. IEEE/ACM Trans Netw 12(3):493–506
7. Maheshwari S, Vasu K, Mahapatra S, Kumar CS (2017) Measurement and analysis of UDP traffic over wi-fi and GPRS.arXiv preprint arXiv:1707.08539
8. Zhao R, Woodford T, Wei T, Qian K, Zhang X (2020, April) M-Cube: a millimeter-wave massive MIMO software radio. In: Proceedings of the 26th annual ınternational conference on mobile computing and networking, pp 1–14
9. Niu Y, Li Y, Jin D, Su L, Vasilakos AV (2015) A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges. Wirel Netw 21(8):2657–2676

# Automatic Vehicle Service Monitoring and Tracking System Using IoT and Machine Learning


**M. S. Srikanth, T. G. Keerthan Kumar, and Vivek Sharma**

**Abstract** Nowadays, vehicle monitoring is emerging as a very tedious job, which requires the maintenance of the record or by recalling the date again and again for the service, and one more problem is tracking the vehicle location for providing better security and safety measures during the travelling. In both cases, it requires more human effort. The proposed model uses the novel technologies like IoT, cloud computing and machine learning. IoT allows various devices to interact and collect data like distance travelled, lubricant level, tyre conditions, smoke emission, other hardware parts conditions and also global positioning system (GPS) to track vehicle location. This data will be collected from various sensors like IR (infrared), MQ 6 sensor, HC SR 04 ultrasonic sensor, light-dependent resistor (LDR) sensor and stored in the cloud storage system. The machine learning algorithm is used to train the proposed model by using the samples data which is collected from the real-time vehicle service stations for service monitoring and GPS data for vehicle tracking purpose. Then this trained model is used to predict the vehicle's condition based on that it will suggest the next date of service. This will help us to condense the quantity of human effort required to predict the vehicle service date. By use of previously fed data and algorithms used to analyse, the model is capable of providing the efficient result. Finally, the collected data is stored in the cloud storage and used to forecast upcoming service date, and all these activities like vehicle service date and GPS location data are provided through an Android application for ease of use to the user and the service provider for their need.

**Keywords** IoT · Machine learning · Android · Cloud computing · GPS · Tracking system · Vehicle monitoring · Sensors

M. S. Srikanth · V. Sharma
Nagarjuna College of Engineering and Technology, Bangalore, India

T. G. K. Kumar (✉)
Siddaganga Institute of Technology, Tumakuru, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_72

## 1 Introduction

In the modern era, everyone is more dependent on the vehicles either by using public or private transportation. The vehicle servicing is a set of the activities that are followed to maintain the vehicle in the good condition. To maintain the vehicle in good condition, the time to time service is essential. The vehicle servicing is all depend on the internal factors like distance travelled, lubricant (oil) level, tyre conditions, the smoke emission, depending on the other hardware parts conditions and external factors like a year of manufacture, vehicle model, driving conditions, the speed of driving, and the weather conditions. For long usage and maintaining its original performance, the vehicle servicing is important for very regular interval.

Traditional vehicle service is carried out in a very random fashion like first whenever the problem occurs while driving, in this case, definitely the owner faces the very big problem in terms of location or time where vehicle gets off while travelling. In another case, when the owner gets the reminder call from the vehicle service station, but in this case, the owner must be aware of the call and he must be ready with all financial required but here, the vehicle owner as to be depend on the service station.

The main problem from traditional vehicle service monitoring is the vehicle owner as to remember the dates of previous service or he or she has to wait for the call from the service station. In such situations, during the travel, if the vehicle may go off or poor performance before the next date of service, vehicle owner totally unaware of what to do next and where to take the vehicle for service. Proposed model uses machine learning which is one of the most trending technologies [1, 2]. As it is proof from the name, it gives the computer system that which makes it more similar to the humans being. Machine learning is actively being used today, in more places than one would expect. One more emerging technology is Internet of things (IoT) [3–5], where all the devices are connected and communicated to each other using various sensors. Finally, the data which user can receive from all connected IoT devices can be stored in the cloud for further processing and analysis [6–8].

## 2 Literature Survey

Many works are carried out previously in the area of Android application development for vehicle service monitoring, but no tracking system. In Anusha et al. [3] proposed a method of contemporary technology by means of Embedded C programming language and the unit developed via LPC2148. Kamijo et al. [9] proposed an Android-based vehicle service status monitoring system in which Android application is used to carry out the vehicle facility user-friendly but this application not concentrated on hardware devices, and it covers few features like notification of the service, service status tracking and monitoring. But here, no concepts of storing the data for future prediction and no synchronization of hardware devices. It means that it is not used any technologies like machine learning and IoT. Pham et al. [10] proposed

vehicle tracking system using GPS and GSM modem by using GPS technology; it is easy to find out the cellular mobile towers and the remote devices and the vehicle location. The GPS technology is used to find out the latitude and longitude in the map where the vehicle is located. But there is no service providing facility is incorporated in this model. So vehicle owner completely unaware of what went wrong in vehicle. Celesti et al. [11] proposed a cloud system with IoT using OpenGTS and MongoDB for monitoring traffic and alert. Mahalle et al. [12] proposed the hybrid system of securing data in the cloud. The hybrid combination is implemented using RSA and AES algorithms to improve the safety of the cloud data. Here, author emphasises on providing security for data which is uploaded. The data which is stored, the downloading of that data is done in such a way that the integrity of data is maintained and the data is retrieved in the secured manner and also the proper usage of the private key, public key and secret keys which involved in encryption and decryption which are the important features of RSA algorithm. Then this encrypted data will be further encrypted with the help of a secret key and public key. By use of this method, it is difficult for any third party to decrypt the actual data easily, and hence, this method provides more security.

## 3 Proposed Framework

In the proposed framework, the vehicle data is captured using various interconnected IoT devices, in order to solve the vehicle owner problem towards monitoring and tracking of vehicle status.

- **The Arduino Uno board**: It is microcontroller board and uses Flash and EEPROM as a storage.
- **NodeMCU**: It is having an USB input and formed by ESP12E. It provides an access to general-purpose input and output (GPIO).
- **Infrared(IR) sensors**: Used for motion detection and used to monitor the heat of the engine.
- **MQ-6 gas sensor**: It is a gas sensor used to detect the gas leakage of a vehicle and its very low-cost device.
- **Ultrasonic sensor**: The accurate distance can be measured by the ultrasonic sensor.
- **Light-dependent resistor (LDR) sensor**: It is made up of semiconductor material and used to sense the light.
- **WIFI module**: Used for WIFI to establish network connectivity across various module.
- **Temperature sensor**: It is used to measure the temperature of any object.
- **Inter-integrated circuit ($I^2C$)**: $I^2C$ used for intention of message between chips exist in on the same printed circuit board.

### *3.1 System Design*

The proposed model is shown in Fig. 1. It consists of three modules, namely data, service side and user side module. Here, the various parameters of a vehicle such as distance travelled, quantity of emitted smoke and lubricant level are collected from various sensors, and data is stored in temporary memory and moved to cloud for future reference using WIFI module by encrypting the data and same data is decrypted and accede by the Android application. The data stored in cloud can be used by multiple linear regression algorithm for prediction of the service date. It uses various parameters like distance travelled, quantity of emitted smoke, lubricant level for training the data model and efficient estimation of the upcoming service date for the vehicle. The flow chart for the Android application is shown in Fig. 2, and both user and service provider can register and login to the application for accessing the data from cloud. In the proposed model, the Android application repeatedly informs the owner about vehicle service date and nearest service provider details using GPS details, and after the vehicle service, a notification will go to vehicle owner regarding service completion and payment for service. Along with this facility, owner can see previous service history also.

Distance is being collected based on the constant speed of the vehicle. The average speed of the vehicle is 25 m/s. So the total distance covered by the vehicle is determined by using Eq. (1).

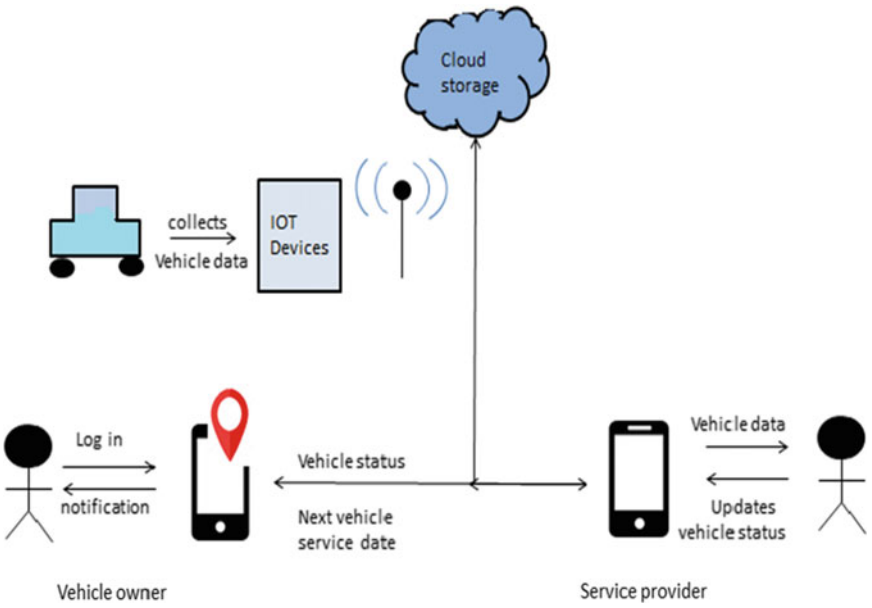



**Fig. 1** Proposed model

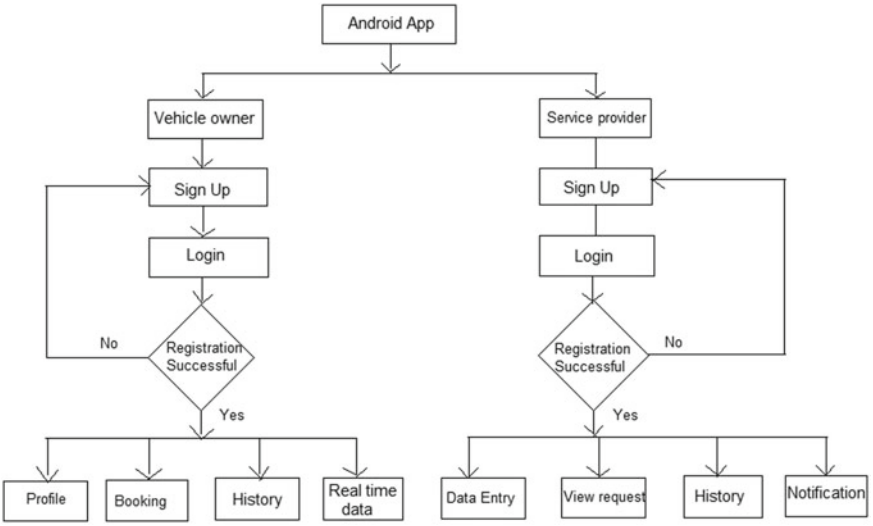



**Fig. 2** Flow chart of application

$$\text{Distance} = \text{Speed (meters)} * \text{Time (in seconds)} \tag{1}$$

The multiple linear regression [13] will consider several prediction parameters for getting the output. The output of this framework is also dependent on several parameters like distance travelled, lubricant level, smoke emission, etc., and against these parameters, the next service date of the vehicle is predicted. These parameters are independent variables and the next service date is the dependent variable. The accuracy of the regression model is being determined by using equation Eq. (2).

$$\text{Co-efficient of determination } \left(R^2\right) = 1 - (\text{SSE/SST}) \tag{2}$$

where

SSE error sum of squares
SST total sum of squares
$R^2$ Co-efficient of determination.

**Multiple Linear Regression Algorithm**

Actual value = regression · predict(X_cordinates).

SS_Residual error = sum((Y-Actual value)ˆ2).

SS_Total = sum((Y-mean(Y))ˆ2).

r_squared = 1 − (float(SS_Residual))/SS_Total.

All the collected data is encrypted using RSA algorithm and stored in cloud storage. When the data is to be stored in the cloud server, the data is encrypted using the public key, and when the data is retrieved using Android application, the private

key is used. This will not let any third party to access the data. It also uses one common key that is being used along with the private key and public key. Hence, a good security is provided.

## *3.2 Implementation*

Fig. 6 shows the architecture model. Here, various sensors like IR, MQ 6, ultrasonic, LDR sensors, etc. are interconnected to the Arduino board [14]. From these sensors, the data is collected in regular interval, and this data will be used for training the model and thereby used for prediction vehicle condition and upcoming service dates. The proposed model consists of three modules, namely data module, service provider module and user (vehicle owner) module.

In the data collection module as shown in Fig. 3, the various real-time data like total distance travelled by the vehicle, GPS location, level of the lubricant, quantity of smoke emitted from the vehicle, all these data are collected from interconnected sensor devices through Arduino board [15] in regular intervals, and finally, data is encrypted using RSA algorithm [16] to protect the data from the attackers and stored in cloud server. Proposed model can use this data for input for machine learning algorithms like multiple linear regression which is used to predict the next date of vehicle service. The next module is at service provider side; through the application, the service provider can view the status of the vehicle which is registered for the service from the vehicle owner. The application retrieves the data from the cloud [17]. After viewing the real-time data service, provider can take the necessary actions for the vehicle and same is stored again in cloud for prediction of the next service

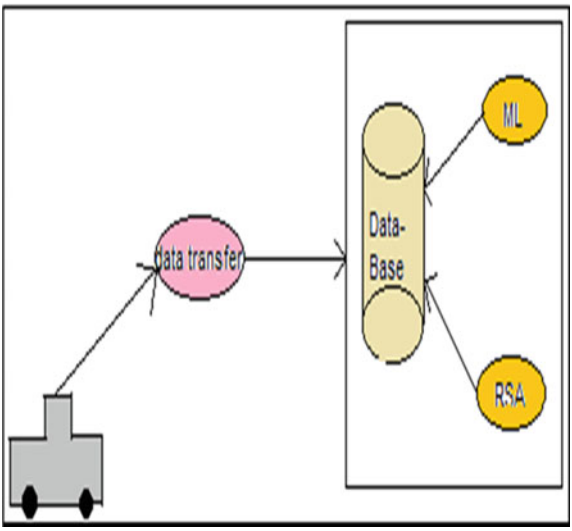



**Fig. 3** Data collection module

date for the vehicle, and in the other side, same can be viewed by the vehicle owner regarding the vehicle service status as well as cost for the service as shown in Fig. 4.

Fig. 5 shows user side module, from which vehicle owner can login to the application then he can update its profile and they can search for the nearest service centre. Once the vehicle reaches the service station for the service, the service provider makes use of this application, and he can view the current status of the vehicle and

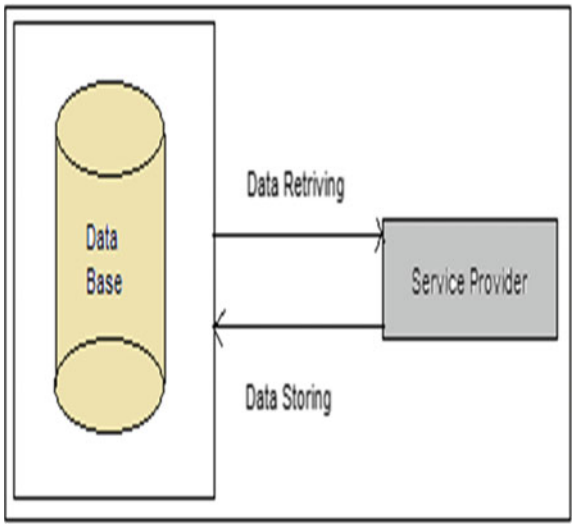



**Fig. 4** Service side module

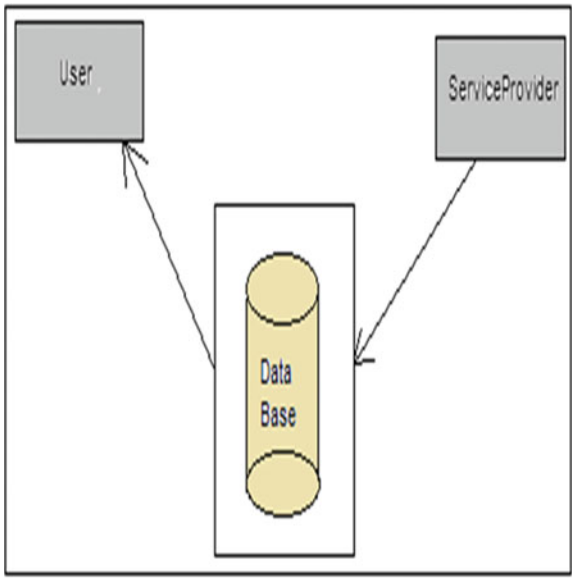



**Fig. 5** User side module

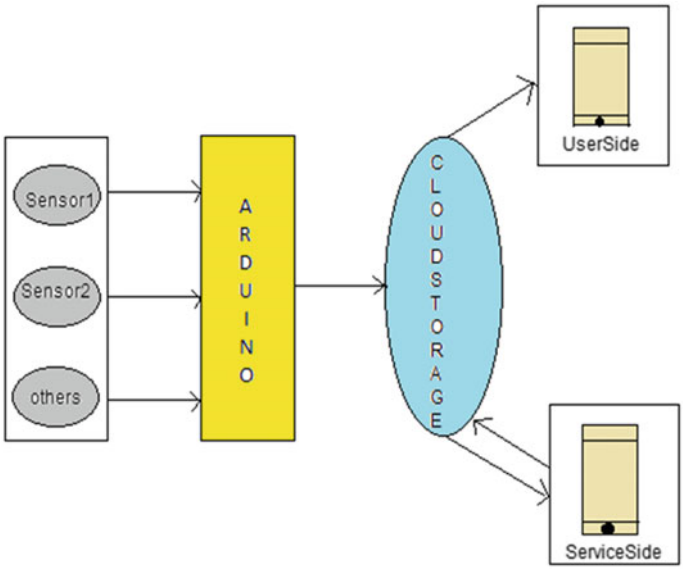



**Fig. 6** Architecture model

do the necessary service, and same will be stored in cloud for future usage by using this data for multiple linear regression algorithm for prediction of next service date as shown in the (Fig. 6).

## 4 Results

Hardware connections to collect distance travelled temperature and level of lubricant are shown in Fig. 7. Hardware set-up for emission test of vehicle is shown in Fig. 8. The use case diagram is represented in Fig. 9. There are two users of this application, one user is a vehicle owner and another one is the service provider. Both vehicle owner and service provider will register for the application and vehicle owner can view the status or current condition of the vehicle after login and service provider can fill the cost for each service and can provide suggestions for the vehicle owner through this application.

The front page and login/sign up page of the Android application shown in Fig. 10a, b, respectively, and Fig. 10c show the registration page of the user. Figure 11a shows the login and sign up/Registration page of the Android application for service provider and Fig. 11b shows the registration page of service provider. The service provider can register by providing service station details like city, vehicle brand and phone number. Figure 12 shows the vehicle booking page; after successful login by the vehicle owner, owner can do vehicle booking by selecting the city where he

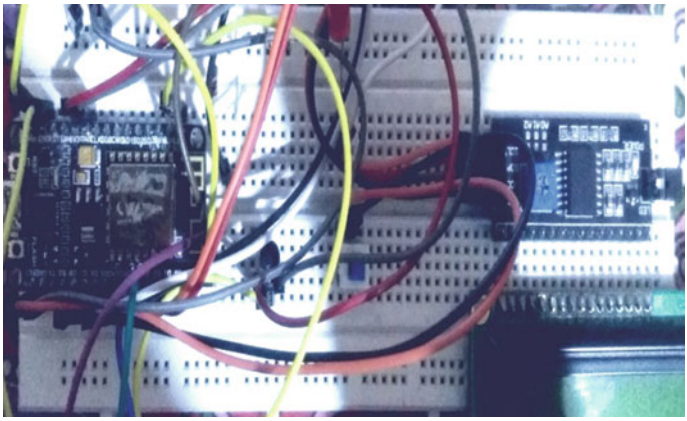


**Fig. 7** Connections to collect distance travelled temperature and level of lubricant

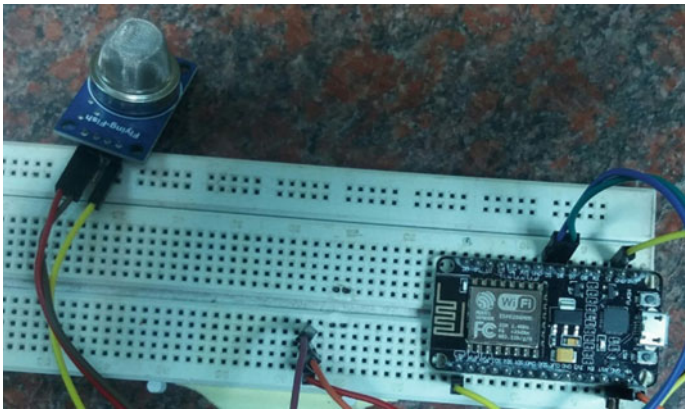



**Fig. 8** Hardware set-up for emission test of vehicle

wants to give it for service, date and vehicle brand type. Service provider can see the registered vehicles owner details for service station by logging in to service login page and can accept or reject the request based on the availability. If the request is accepted, then the following factors are checked based on the previous vehicle service history and present status of various factors like battery, brake cam, brake effectiveness, carburettor assembly, carburettor breath hose, clutch, control cables, engine, front wheel, headlamp, kick starter, oil, sparkplug, steering, transmission oil, which are obtained from the various supporting sensors used in this model and are passed as inputs for the multiple linear regression and other relevant algorithms to carry out the service. Once the service is done, same is updated in Android

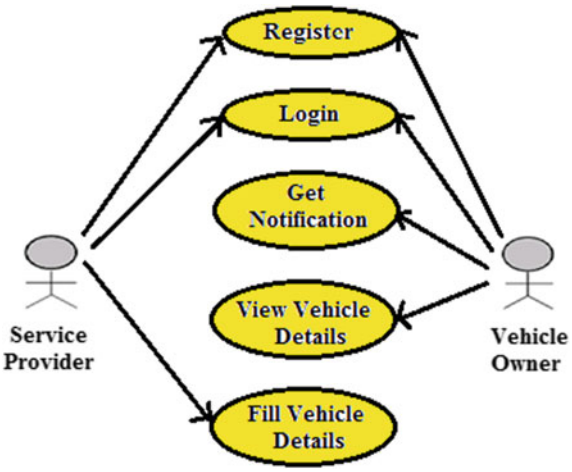



**Fig. 9** Use case diagram

application along with cost of each and same thing is reflected on vehicle owner page. Figure 13 shows the coefficients of multiple linear regression algorithm. By using multiple linear regression algorithm, next service details can be predicted.

## 5 Conclusions

This paper has successfully proposed an Android application for providing vehicle service monitoring and tracking more efficiently to the user. The IoT technology helps in fast gathering of data from various vehicle components and machine learning algorithms are used to process the collected data and give best results to the users. GPS location helps to keep track of vehicle and providing needs in case of emergency. The proposed system helps in automatic monitoring and servicing of the vehicle by notifying the service dates from time to time to the vehicle owner. By this application, the vehicle owner will be conscious of the present status of the vehicle which implemented using IoT and machine learning algorithms. The service provider also can make use of this application for analysing the vehicle status and for prediction of service cost in early time, and user also gets to know about complete service details of his/her vehicle from the application itself. This model helps to keep vehicle in good condition. Finally, this model overcomes the limitations and risks facing the traditional methods and gives the efficient service.

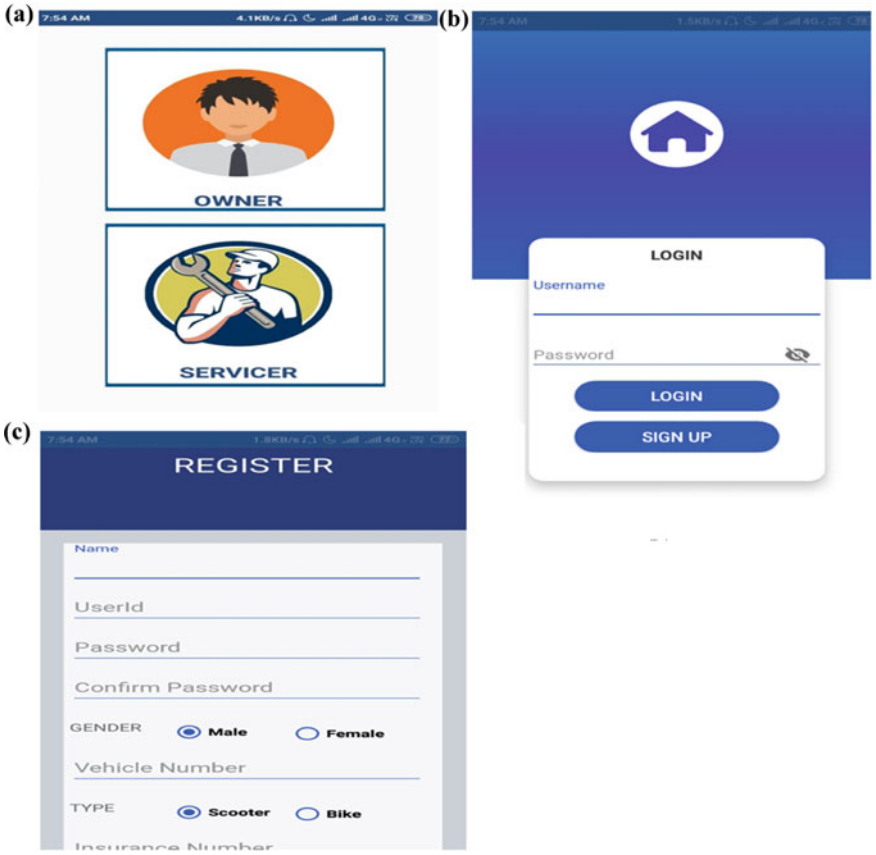



**Fig. 10** Android application **a** front page. **b** Login/sign up page of vehicle owner. **c** Registration page of owner

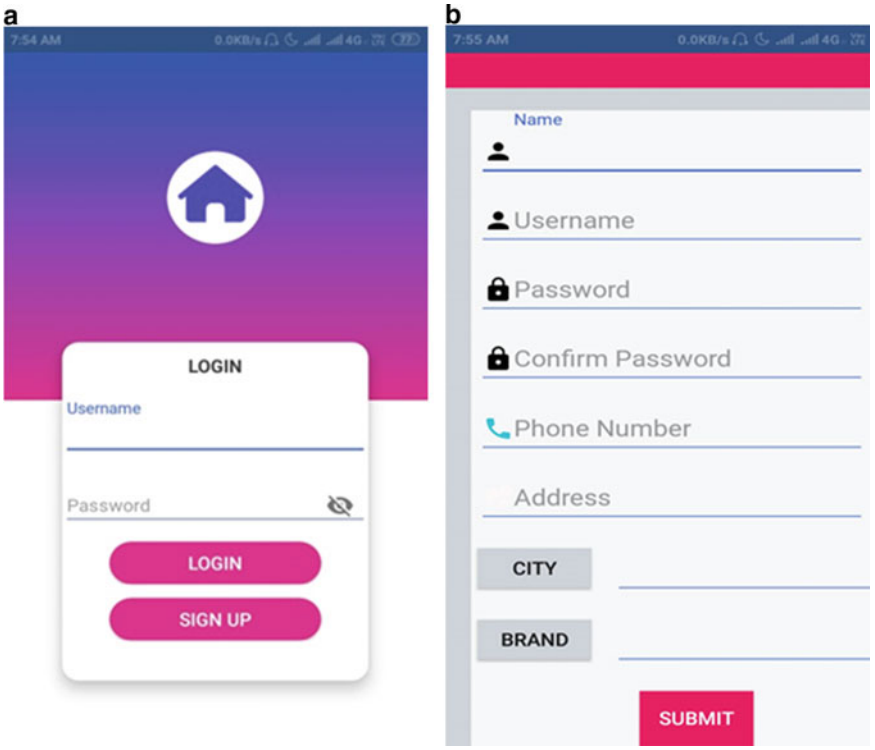



**Fig. 11** Android application. **a** Login/sign up page of service provider. **b** Registration page of service provider

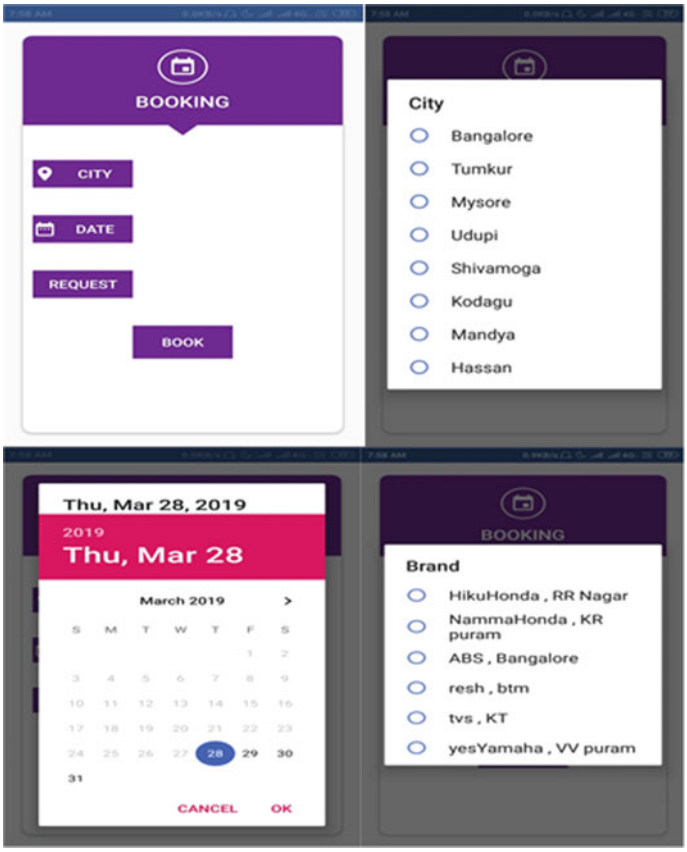



**Fig. 12** Vehicle booking page

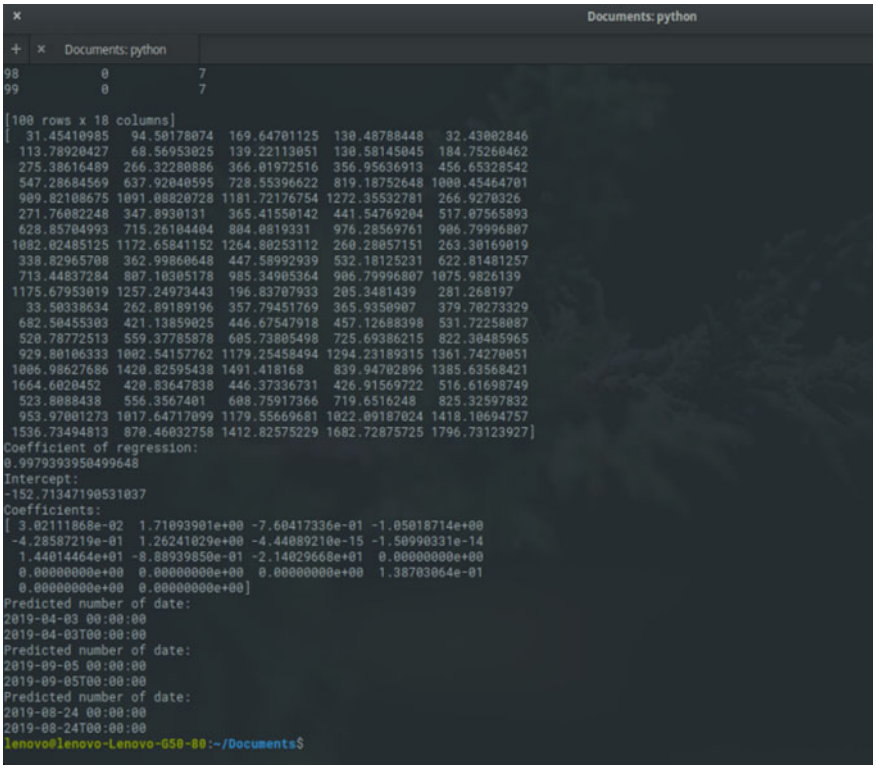



**Fig. 13** Coefficients of multiple linear regression algorithm

## References


1. Keerthan Kumar TG, Shubha C, Sushma SA (2019) Random forest algorithm for soil fertility prediction and grading using machine learning. Int J Innov Technol Explor Eng 9(1). https://doi.org/10.35940/ijitee.L3609.119119
2. Pan L, Zhang X, Liu J (2019) A comparison of three widely used GPS triple-frequency precise point positioning models. GPS Solut 23:121. https://doi.org/10.1007/s10291-019-0914-3
3. Anusha A, Ahmed SM (2017) Vehicle tracking and monitoring system to enhance the safety and security driving using IoT. In: Proceedings of international conference on recent trends in electrical, electronics and computing technologies, pp 49–53
4. Stergiou C, Psannis KE, Kim B-G, Gupta B (2018) Secure integration of IoT and Cloud computing. Future Gener Comput Syst 78(3):964–975
5. Chen JIZ, Lai K-L (2020) Machine learning based energy management at internet of things network nodes. J: J Trends Comput Sci Smart Technol (3):127–133
6. Kajol R, Akshay KK, Keerthan Kumar TG (2018) Automated agricultural field analysis and monitoring system using IOT. Int J Inf Eng Electron Bus 10(2):17 (Hong Kong). https://doi.org/10.5815/ijieeb.2018.02.03
7. Keerthan Kumar TG, Virupakshaiah HK, Nanda KV (2016) Ensuring an online Chat mechanism with accountability to sharing the non-downloadable file from the Cloud. In: Proceedings of 2nd international conference on applied and theoretical computing and communication technology, pp 718–721

8. Wei W, Li Y; Wang X; Liu J, Zhang X (2018) Detecting Android malicious apps and categorizing benign apps with ensemble of classifiers. Future Gener Comput Syst 78(3): 987–994
9. Kamijo S, Matsushita Y, Ikeuchi K, Sakauchi M (2000) Traffic monitoring and accident detection at intersections. IEEE Trans Intell Transp Syst 1(2):108–118
10. Liu X, Jiang W, Chen H et al (2019) An analysis of inter-system biases in BDS/GPS precise point positioning. GPS Solut 23:116. https://doi.org/10.1007/s10291-019-0906-3
11. Celesti A, Galletta A, Carnevale L, Fazio M, Ĺay-Ekuakille A, Villari M (2018) An IoT Cloud system for traffic monitoring and vehicular accidents prevention based on mobile sensor data processing. IEEE Sens J 18(12):4795–4802
12. Mahalle VS, Shahade AK (2014) Enhancing the data security in Cloud by implementing hybrid (Rsa & Aes) encryption algorithm. In 2014 International Conference on Power, Automation and Communication (INPAC), Amravati, India, 2014, pp. 146–149. https://doi.org/10.1109/INPAC.2014.6981152
13. Yan X, Xie H, Tong W (2011) A multiple linear regression data predicting method using correlation analysis for wireless sensor networks. In: Proceedings of 2011 cross strait quad-regional radio science and wireless technology conference, pp 960–963
14. Vimos V, Cabrera EJS (2018) Results of the implementation of a sensor network based on Arduino devices and multiplatform applications using the standard OPC UA. IEEE Latin Am Trans 16(9):2496–2502
15. Wang K, Hou Y, Xu Y (2016) Design and implementation of remote control system between Android platform. In: Proceedings of international conference on information system and artificial intelligence, pp 143–147
16. Sun H, Wu M, Ting W, Hinek MJ (2007) Dual RSA and its security analysis. IEEE Trans Inf Theor 53(8):2922–2933
17. Seo D, Kim S, Song G (2017) Mutual exclusion method in client-side aggregation of cloud storage. IEEE Trans Consum Electron 63(2):185–190

# Machine Learning-Based Application to Detect Pepper Leaf Diseases Using HistGradientBoosting Classifier with Fused HOG and LBP Features


**Matta Bharathi Devi and K. Amarendra**

**Abstract** Pepper leaf disease detection is one of the interesting challenges in the field of machine learning. In this paper, a machine learning-based approach is proposed to extract texture features and use dimensionality reduction techniques called principal component analysis (PCA) and create a composite feature descriptor. There are two different texture-based feature representations extracted by using HOG and LBP feature engineering techniques were used for the pepper leaf images, and PCA is applied to obtain reduced representations. These representations are fused and passed to machine learning models like logistic regression, naïve Bayes, decision tree, support vector machine, and HistGradientBoosting classifier for classification. HistGradientBoosting classifier achieved highest the accuracy of 89.11% and outperformed other models.

**Keywords** Histogram of oriented gradients (HOG) · Local binary pattern (LBP) · Principal component analysis (PCA) · HistGradientBoosting classifier (HGB) · Machine learning

## 1 Introduction

Detecting plant leaf diseases is one of the major challenges faced by farmers in agriculture. It is very important to identify the type of leaf diseases accurately for the appropriate use of pesticides. Any mistakes in identifying diseases of plants lead to reduced yield. Plant diseases can be either biotic [1, 2] or abiotic. The primary cause behind biotic diseases is various living organisms like bacteria, viruses, and fungi. Biotic diseases are affected by viruses, unlike abiotic diseases which are affected by inorganic conditions like weather changes and chemicals. Identifying leaf diseases accurately by observing with the naked eye is a difficult task. Hence,

M. B. Devi (✉) · K. Amarendra
Department of CSE, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India

K. Amarendra
e-mail: amarendra@kluniversity.in



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_73

there is a requirement for an application that can detect leaf diseases accurately. There are various automated applications to identify plant leaf diseases. Most of them used texture representations extracted from leaf images with conventional machine learning models [3–5].

Most of the recent works in the literature used feature extraction techniques like histogram of oriented gradients (HOG), local binary patterns (LBP), gray-level co-occurrence matrix (GLCM) are used in the literature to extract texture-based features from plant leaf images [6, 7]. These features were provided to popular classifiers like support vector machine (SVM) to categorize different types of diseased leaves [8, 9]. However, when using these features directly along with ML models results in reduced performance, so, in this work, the investigation is made to reduce the dimensions of texture features and blend them to get composite representation with pepper leaf dataset [10–12].

Initially, the preprocessing is performed to remove background noise obtained during image acquisition. Later, the two types of texture-based features extracted from pepper leaf images are HOG and LBP feature engineering methods and the principal component analysis (PCA) dimensionality reduction technique is applied to get reduced representations of HOG and LBP features. In our experiments, it can be observed that HOG features are better than LBP. Using reduced representations leads to improve performance. When LBP features are fused with HOG, composite representations are obtained. These representations contain more discriminant information which helps classification models to identify pepper leaf diseases accurately. Our proposed fused representation achieved the highest accuracy of 89.11% with the HGB Classifier.

## 2 Related Work

This part of the paper provides an overview of various methodologies employed for detecting plant leaf diseases in past. The first part of this section describes various preprocessing techniques used in the literature followed by feature engineering methods algorithms that are used for classification.

In the recent past, several preprocessing techniques have been applied to plant leaf images to correctly identify the type of plant diseases. Most of the previous works used image processing techniques and applied them to smooth, sharpening filters enhance the image and used several filters to remove additive noise from the images [13, 14]. ROI segmentation is a major task employed to detect and segment diseased portions from images to improve the performance of automated plant leaf disease diagnosis systems [15–17].

Texture-based features obtained from images play a vital role and affect the performance of image classification systems. Histogram of oriented gradients (HOG), GIST, scale-invariant feature transform (SIFT), local binary patterns (LBP) are majorly employed feature engineering algorithms to obtain intensity and texture-based features [6, 18, 19]. These feature engineering methods are employed in various tasks like medical image classification, scene classification, object recognition, and

leaf disease identification [20–23]. Most machine learning algorithms like K-nearest neighbor classifier (KNN), decision tree classifier, random forest, support vector machine (SVM), naïve Bayes classifier are trained on these texture-based features for classification purpose [8].

A K-nearest neighbor (KNN) classifier with gray-level co-occurrence matrix (GLCM) texture features of plant leaf images was used to identify plant leaf diseases [24, 25]. Another machine learning-based system is was proposed for grapes plant leaf disease detection by Harshal Waghmare et al. At first, the background of all images was removed and segmentation is performed as a preprocessing step. A high-pass filter is applied on segmented images to analyze the diseased part of the leaf. Local binary patterns (LBP)-based texture features are extracted from preprocessed images, and these features were used to identify different types of grape plant diseases using support vector machine (SVM) classifier [20, 26]. A cotton leaf disease detection and classification technique based on machine learning and image processing tools are proposed by Pooja et al. Initially, region of interest (ROI) is segmented from plant leaf images using image processing tools and features are extracted. These features were passed to SVM classifier to identify the type of disease [27, 28].

In this work, the HOG and LBP features extracted from pepper leaf images and fuse them to create a composite representation. These representations are then projected to a lower dimension using PCA. Then apply various popular classification algorithms like logistic regression, naïve Bayes classifier, decision tree classifier, support vector machine (SVM) with linear and radial basis function (RBF) kernel, and HistGradientBoosting classifier for classification purpose.

## 3 Proposed Methodology

This part of the paper illustrates various stages of the proposed method for pepper leaf disease detection. Our proposed work consists of four phases, followed one after the other. They are data preprocessing, feature extraction, dimensionality reduction, and classification.

### *3.1 Data Preprocessing*

Data acquired from real-world consist of random noise in the background. So, background subtraction is performed on pepper leaf images to remove random background noise. This is done by creating a suitable mask for every image present in the dataset, and then background removal operation is performed by using corresponding masks. Figure 1a represents images from the original dataset, and Fig. 1b represents background removed images. These processed images are passed to the feature extraction phase.

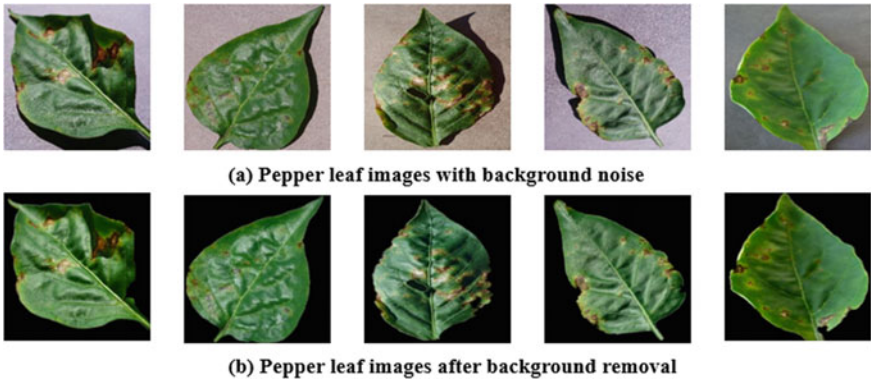



**Fig. 1** Pepper leaf images before and after preprocessing

## *3.2 Feature Extraction*

Feature extraction is an important phase in any machine learning task. In our work, the two different feature extraction techniques chosen are histogram of oriented gradients (HOG) and local binary patterns (LBP) which extracts texture-based features from pepper leaf images.

### 3.2.1 Feature Extraction from Pepper Leaf Images Using HOG

The HOG feature descriptor counts the occurrences of gradient orientation in localized portions of an image. Initially, all processed images of dimension (256 × 256) are reshaped to (64 × 128) dimensions. Next, changes in X- and Y-directions of images (gradients) are computed by dividing the entire image into (8 × 8) patches. Next, magnitude and orientations are computed by using gradients. Then, a histogram of gradients is calculated for each (8 × 8) cell and these cells are combined to create (16 × 16) cells. The gradients of these cells are normalized to get a vector of (1 × 36) dimensions for each cell. Finally, for every image of dimension (64 × 128) obtained a feature vector of 3780 dimensions. This feature descriptor is normalized using the min–max normalization method. Figure 2 represents a histogram of oriented gradients computer for a given pepper leaf image.

### 3.2.2 Feature Extraction from Pepper Leaf Images Using LBP

Local binary patterns (LBP) compute texture features from local regions instead of computing global texture features as in the case of gray-level co-occurrence matrix (GLCM). Initially, all processed images of dimension (256 × 256) are reshaped to (128 × 128) dimensions. Next, all these images are converted to gray scale. LBP

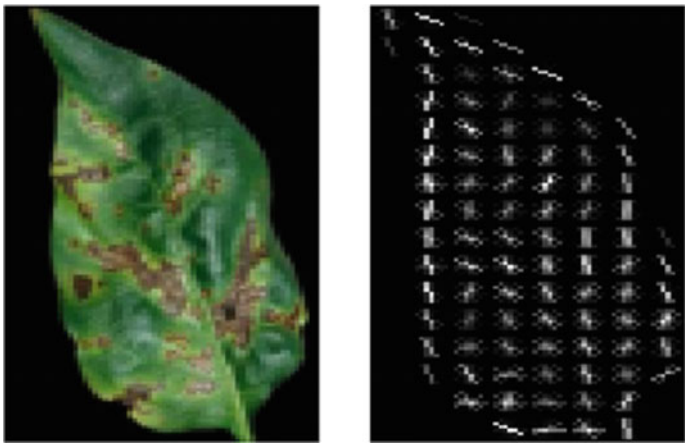


**Fig. 2** Histogram of gradients for a given input image

histogram is obtained from those images by appropriately selecting p and r values, where p represents the number of points in neighborhood of a pixel and r is the radios. Finally, for every image of dimension (128 × 128) obtained a feature vector of 26 dimensions. This feature descriptor is normalized using the min–max normalization method. OpenCV module of python is used to extract LBP features.

### *3.3 Dimensionality Reduction*

In this phase, all the features of dimension 3780 obtained from the HOG feature extraction technique and 26 dimensions obtained from LBP are projected into lower dimensional space with 512 and 13 dimensions for HOG and LBP, respectively. For this, the principal component analysis is used. In the case of limited data, high-dimensional features may lead to the affliction of dimensionality. To resolve this

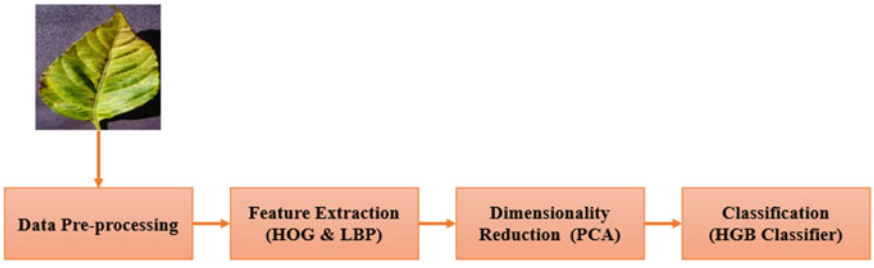



**Fig. 3** Architecture of proposed system for pepper leaf disease detection

problem, this module is included in our work. Figure 3 represents the architecture of the proposed method for pepper leaf disease detection.

### *3.4 Classification*

In this work, popular classification models like logistic regression, naïve Bayes classifier, decision tree classifier, support vector machine (SVM) with linear and radial basis function (RBF) kernel, and HistGradientBoosting classifier were utilized. The performance extracted features are tested before and after applying the dimensionality reduction. The SVM along with the RBF kernel and HistGradientBoosting kernels performs better than other classifiers for both HOG and LBP features in both the cases of dimensionality reduction. Finally, the HOG and LBP features were fused to form a composite feature representation of dimension 3806. With these features, HistGradientBoosting classifiers achieved the highest accuracy of 89.11% and outperformed all other models.

## 4 Experimental Results

This section provides a precise picture of the experiments conducted and the results obtained using the proposed method. Initially, an overview of the dataset used for experiments is described followed by evaluation metrics used to measure the performance of the proposed method. Finally, a summary of experiments and their results is provided.

### *4.1 Pepper Leaf Disease Dataset*

The pepper leaf disease dataset is used, and the part of the plant village dataset is also used which contains 54,306 samples of 26 types of diseased leaf images belonging to 14 types of plant species. This dataset contains 2475 samples of pepper plants representing both healthy and diseased leaves. 997 samples belong to the healthy category and 1478 samples belong to the bacterial spot category. Totally 1980 samples are considered for training, and 495 samples are used for testing model performance.

**Table 1** Performance of ML algorithms with HOG features

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| LR | 75.4 | 75 | 75 | 75 |
| Naïve Bayes | 70.97 | 78 | 71 | 66 |
| Decision tree | 67.74 | 67 | 68 | 67 |
| SVM—linear | 77.82 | 78 | 78 | 78 |
| SVM—RBF | 83.06 | 84 | 83 | 83 |
| HGB | 84.81 | 85 | 84 | 85 |

## *4.2 Performance Evaluation Measures*

Different classification model performance evaluation measures like accuracy, precision, recall, F1-score are calculated to prove the efficiency of the proposed method on test data. These measures can be computed using the confusion matrix.

## *4.3 Result Analysis*

The experiments were conducted in three different ways to check the performance of classification models with HOG, LBP, and fused features before and after applying PCA for pepper leaf disease detection.

### 4.3.1 Experiments Without Dimensionality Reduction

These experiments are conducted to check how well HOG and LBP features can detect pepper leaf diseases before dimensionality reduction.

From Table 1, it is clear that HistGradientBoosting classifier outperformed other ML by achieving an accuracy of 84.81%. Comparatively, the decision tree classifier could not perform well. SVM with RBF kernel also obtained an accuracy of 83.06% which is the second highest measure. From Table 2, it can be observed that the HGB classifier achieved 83.87% accuracy with LBP features and the naïve Bayes classifier obtained lower accuracy. From both experiments, the HOG features perform better than LBP features for the task of pepper leaf disease detection, and the HGB classifier outperformed all other models used, with both types of features.

### 4.3.2 Experiments After Applying Dimensionality Reduction

These experiments are conducted to check how well HOG and LBP features can detect pepper leaf diseases after dimensionality reduction (Tables 3 and 4).

**Table 2** Performance of ML algorithms with LBP features

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| LR | 80.24 | 80 | 80 | 80 |
| Naïve Bayes | 67.74 | 69 | 68 | 68 |
| Decision tree | 72.58 | 73 | 73 | 73 |
| SVM—linear | 81.05 | 81 | 81 | 81 |
| SVM—RBF | 81.85 | 82 | 82 | 82 |
| HGB | 83.87 | 84 | 84 | 84 |

**Table 3** Performance of ML algorithms with HOG features

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| LR | 73.39 | 73 | 73 | 73 |
| Naïve Bayes | 76.21 | 76 | 76 | 76 |
| Decision tree | 62.5 | 62 | 62 | 62 |
| SVM—linear | 77.98 | 79 | 78 | 79 |
| SVM—RBF | 84.27 | 84 | 84 | 84 |
| HGB | 85.47 | 85 | 85 | 85 |

**Table 4** Performance of ML algorithms with LBP features

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| LR | 79.44 | 79 | 79 | 79 |
| Naïve Bayes | 82.26 | 82 | 82 | 82 |
| Decision tree | 77.02 | 77 | 77 | 77 |
| SVM—linear | 80.24 | 80 | 80 | 80 |
| SVM—RBF | 83.47 | 83 | 83 | 83 |
| HGB | 84.27 | 84 | 84 | 84 |

From previous experiments, it is clear that after applying PCA, there is a significant improvement in the performance of classification models with both, HOG and LBP features. HGB classifier followed the same trend and outperformed other classification models with both HOG and LBP features after applying PCA. Even after reduced dimension, there is a significant improvement in all measures used to test the efficiency of models.

### 4.3.3 Experiments with Fused Features of HOG and LBP

This experiment is conducted to check the performance of ML models with composite representation obtained after blending LBP features with HOG features.

**Table 5** Performance of ML algorithms with fused HOG and LBP features

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| LR | 80.24 | 80 | 80 | 80 |
| Naïve Bayes | 75.81 | 76 | 76 | 76 |
| Decision tree | 69.35 | 70 | 69 | 69 |
| SVM—linear | 81.05 | 81 | 81 | 81 |
| SVM—RBF | 88.71 | 89 | 89 | 89 |
| HGB | 89.11 | 89 | 89 | 89 |

From Table 5, it is clear that the HGB classifier trained on fused feature descriptor obtained 89.11% accuracy which is highest when compared with the performance of the same classifier trained on HOG and LBP features before and after applying PCA. So, it can be concluded that fused texture representations of pepper leaf images help to identify diseases accurately rather than using conventional usage of LBP and HOG features.

## 5 Conclusion

Pepper is the most recently used ingredient in dishes. Identifying pepper leaf diseases is a challenge for farmers. There is a high requirement to automate the process of detecting pepper leaf diseased for correct usage of pesticides and reduce the loss of yield. In this paper, the performance of various classification models along with two different types texture-based features was investigated. During our experiments, it can be observed that models can perform well with reduced representations of HOG and LBP features rather than using them directly. The fused representation of HOG and LBP features helped the models to perform well, and there is a 4% improvement in accuracy with fused features. In our experiments, it can also be observed that the HGB classifier outperforms other ML algorithms in every case.

## References

1. Husin ZB, Aziz AHBA, Shakaff AYBM, Farook RBSM (2012) Feasibility study on plant chili disease detection using image processing techniques. In: IEEE 3rd international conference on intelligent system modeling and simulation ISMS, Kota Kinabalu, pp 291–296
2. Balakrishna G, Moparthi NR (2020) Study report on indian agriculture with IoT. Int J Electr Comput Eng 10(3):2322
3. Barbedo JGA (2016) A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst Eng 144:52–60
4. Mannepalli K, Sastry PN, Suman M (2017) Accent recognition system using deep belief networks for Telugu speech signals. Int J Speech Technol 19(1):87–93

5. Rajesh Kumar T, Suresh GR, Kanaga Suba Raja S (2018) Conversion of non audible murmur to normal speech based on full-rank Gaussian Mixture model. J Comput Theoretical NanoSci 15(1):185–190
6. Mahapatra S, Kannoth S, Chiliveri R, Dhannawat R (2020) Plant leaf classification and disease recognition using SVM, a machine learning approach. Sustain Humanosphere 16(1):1817–1825
7. Ayushree, Balaji GN (2018) Comparative analysis of Coherent routing using machine learning approach in MANET. Smart Comput Inform 731–741
8. Bhagat M, Kumar D, Haque I, Munda HS, Bhagat R (2020) Plant leaf disease classification using grid search based SVM. In: 2nd international conference on data, engineering and applications (IDEA). IEEE, pp 1–6
9. Puri GD, Haritha D (2018) Framework to avoid similarity attack in big streaming data. Int J Electr Comput Eng 8(5):2920–2925
10. Anjali Devi S, Siva Kumar S (2018) Comprehensive survey on sentiment analysis based on workflow foundation. J Adv Res Dyn Control Syst 10(9 Special Issue):1189–1120
11. Rajesh Kumar T, Vamsidhar T, Harika B, Madan Kumar T, Nissy R (2019) Students performance prediction using data mining techniques. IEEE Explorer (ICISS-2019). 978-1-5386-7798-8
12. Talasila V, Rajesh Kumar T, Sai CP, Satya Sai S, Ayyappa (2019) Predicting the risk of heart failure with EHR sequential data modelling. Int J Recent Technol Eng (IJRTE) 6(7):458–461. 2277-3878
13. Asfarian A, Herdiyeni Y, Rauf A, Mutaqin KM (2013) Paddy diseases identification with texture analysis using fractal descriptors based on Fourier spectrum. In: IEEE international conference on computer, control, informatics and its applications IC3INA, Jakarta, pp 77–81
14. Bommadevara HSA, Sowmya Y, Pradeepini G (2019) Heart disease prediction using machine learning algorithms. Int J Innov Technol Exploring Eng 8(5):270–272
15. Khirade SD, Patil AB (2015) Plant disease detection using image processing. In: IEEE international conference on computing communication control and automation (ICCUBEA), pp 768–771
16. Rajesh Kumar T, Suresh GR, Kanaga Subaraja S, Karthikeyan C (2020) Taylor-AMS features and deep convolutional neural network for converting non-audible murmur to normal speech. Comput Intell 1–24
17. Dudi B, Rajesh V (2018) An efficient algorithm for medicinal plant recognition. Int J Pharm Res 10(3):87–93
18. Patil R, Kumar S (2020) Bibliometric survey on diagnosis of plant leaf diseases using artificial intelligence. Int J Mod Agric 9(3):1111–1131
19. Dudi B, Rajesh V (2019) Medicinal plant recognition based on cnn and machine learning. Int J Adv Trends Comput Sci Eng 8(4):628–631
20. Bodapati JD, Veeranjaneyulu N, Shareef SN, Hakak S, Bilal M, Maddikunta PKR, Jo O (2020) Blended multi-modal deep ConvNet features for diabetic retinopathy severity prediction. Electronics 9(6):914
21. Dondeti, V, Bodapati JD, Shareef SN, Naralasetti V (2020) Deep convolution features in non-linear embedding space for fundus image classification deep convolution features in non-linear embedding space for fundus image classification
22. Bodapati JD, Shaik NS, Naralasetti V, Mundukur NB (2020) Joint training of two-channel deep neural network for brain tumor classification. Signal Image Video Process 1–8
23. Bodapati JD, Veeranjaneyulu N, Shaik S (2019) Sentiment analysis from movie reviews using LSTMs. Ingénierie Des Systèmes D Inf 24(1):125–129
24. Trivedi J, Shamnani Y, Gajjar R (2020) Plant leaf disease detection using machine learning. In: International conference on emerging technology trends in electronics communication and networking. Springer, Singapore, pp 267–276
25. Inthiyaz S, Prasad MVD, Lakshmi RUS, Sai NS, Kumar PP, Ahammad SH (2019) Agriculture based plant leaf health assessment tool: a deep learning perspective. Int J Emerg Trends Eng Res 7(11):690–694

26. Anila M, Pradeepini G (2017) Study of prediction algorithms for selecting appropriate classifier in machine learning. J Adv Res Dyn Control Syst 9(Special Issue 18):257–268
27. Waghmare H, Kokare R, Dandawate Y (2016, February) Detection and classification of diseases of grape plant using opposite colour local binary pattern feature and machine learning for automated decision support system. In: 2016 3rd international conference on signal processing and integrated networks (SPIN). IEEE, pp 513–518
28. Shariff MN, Saisambasivarao B, Vishvak T, Rajesh Kumar T (2017) Biometric user identity verification using speech recognition based on ANN/HMM. J Adv Res Dyn Control Syst 9(12 Special issue):1739–1748

# Efficacy of Indian Government Welfare Schemes Using Aspect-Based Sentimental Analysis


**Maninder Kaur, Akshay Girdhar, and Inderjeet Singh**

**Abstract** One of the simplest methods to understand people's thoughts using images or text is commonly given as sentiment analysis. Sentiment analysis is used mostly in products advertisement and promotion depends on the user's opinion. The process is based on the aspect-based sentiment analysis and it is used to understand and find out what someone is speaking about, and likeness and dislikeness. One of the real-world models of the perfect realm of this subject is the huge number of available Indian welfare plans like Swachh Bharat Abhiyan and Jan Dhan Yojna. In this paper, labeled data is used on the basis of polarity. Tweets are preprocessed and unigram features are then extracted. In the initial steps, tokenization process, stop word removal process, and stemming process are performed as preprocessing to remove duplicate data. The unigram features and labels trained by support vector machine (SVM), K-nearest neighbor (KNN), and a combination of SVM, KNN, and random forest as a proposed model are used in the presented work. Implementation of experimental proposed approach demonstrates that better results in accuracy and precision than SVM and KNN.

**Keywords** Sentiment analysis · Aspect · Support vector machine · K-nearest neighbor

M. Kaur (✉)
Department of Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana, India

A. Girdhar
Department of Information Technology, Guru Nanak Dev Engineering College, Ludhiana, India
e-mail: akshay_girdhar@gndec.ac.in

I. Singh
Information Technology, Govt. Polytechnic College, Bathinda, India



A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7_74

## 1 Introduction

In natural language processing, sentiment analysis is considered as most significant tool because it opens up numerous possibilities to understand people's sentiments on different topics. The purpose of an aspect-based sentiment analysis is to detect the features of the particular entity. The positive and negative aspects of a particular topic can be analyzed through aspect-based sentiment analysis [1]. In this research paper, aspect-based sentiment analysis (ASBA) is implemented on the tweets of the government welfare schemes. This type of analysis is mainly domain specific. The government has launched the various welfare schemes for schools, states, as well as center. The given outlines progress in collaboration with center and state governments. The welfare schemes have been mostly introduced to develop the weaker and minority section of the society [2]. Many schemes are launched but in this research work is on Swachh Bharat Abhiyan and Jan Dhan Yojna. These schemes empower every Indian by helping them financially and providing the basic facilities [3].

A. *Two Types of Sentiment Analysis*

Document Level: If the analysis is performed using documents to identify the positive and negative view of the single entity, then it is document level analysis [4].

Comparative: In many cases, users express their views by comparing with the similar product or entity. The main goal is to identify the opinion from the comparative sentence [4].

The research work is structured as follows: Related works are presented in Sect. 2, proposed model is presented in Sect. 3, results and its discussions are presented in Sect. 4, and finally, conclusion is presented in Sect. 5.

## 2 Related Work

Various approaches proposed by the researchers for sentiment analysis are presented in this section. Shidaganti et al. [5] presented a technique that used data mining along with machine learning as a combination process. Clustering is done by using k-means clustering and hierarchical clustering approach. This approach helps to analyze the data of different organizations which helps in understanding the thoughts and opinions related to the product. Rout et al. [6] presented the way to deal with unstructured data of social media like blogs, Twitter for sentiment and emotion analysis. The supervised and unsupervised approaches are performed on different databases. The unsupervised approach was used for automatic identification of sentiments for the tweets. The sentiment identification is done by using maximum entropy, multinomial Naïve Bayes, and support vector machine classifier. This approach also works well if it will apply on the larger dataset in future. Mumtaz et al. [7] presented an approach which is a combination of lexical-based approach and machine learning. In this proposed work, hybrid approach was used, which gave high accuracy than

the classical lexical method and provides the enhanced redundancy than learning approach. Natural language processing to extract sentiments from texts is performed in the approach. Al-Smadi et al. [8] worked on aspect-based sentiment analysis to review the hotels in Arabic. Long short-term memory (LSTM) neural networks are used in the research model and it was implemented in two levels that are character level and aspect based with random field classifier and polarity classification based. Chen et al. [9] presented a visualization approach called Tag Net which is used for sentiment analysis. This approach combines the improved node-link diagrams with tag clouds to obtain heterogeneous time varying information. Large datasets scalability is improved using the proposed algorithm. Fouad et al. [10] presented a model for Twitter sentiment analysis which describes the tweet is positive or negative by using the concept of machine learning. The presented work uses different methods to label the input in the training phase using different datasets. The classification is also done by using different classifiers to compare their performances. The concept of feature selection and information gain is used in this work. Jones et al. [11] focused on the implication of PMJDY scheme for the development all over India. The main aim of scheme was to make India digitized, even the rural areas are conscious about their bank accounts to obtain government offering benefits directly. Demonetization of currency and all these measures lead toward the progress in the structure of Indian economy. This scheme ensures the better quality of life in the country and helps improve the living status of the people. India is the only nation where 50% of people are in the working age group. Khan et al. [12] focused on the challenges that Twitter information faces, concentrating on order issues, and afterward consider these streams for supposition mining and sentiment analysis. To manage gushing unequal classes, the sliding window Kappa measurement for assessment in time-changing information streams was used. Utilizing this measurement an investigation is performed on Twitter information for utilizing learning calculations for information streams. Greica et al. [13] presented an analytic model which performs aspect classification, separation and polarity classification for analysis. The testing of domain aspect and sentiment classification is performed on the multilingual SemEval dataset. Pham et al. [14] presented the multilayer architecture for representing customer review. It extracts the views for product from the sentiment aspects and sentences related to review. The multilayer architecture represents the different level of sentiments for input text. This model is integrated with neural network for prediction of overall ratings of the product. Rathan et al. [15] worked on the review of mobile phones by using the tweets. In this lexicon-based approach is used for data labeling and this improves the classification process. The support vector machine classifier classifies the tweets with efficient accuracy level. Kim et al. [16] worked on the co-occurrence of data by using supervised and unsupervised methods. A framework is used for processing texts reviews. This work also finds the aspect categories according to review sentences and provides effective outcomes from the F-score. Pannala et al. [17] presented the existing work for the opinion mining that was done on the word level, not on the sentence level. The work was done on the trained dataset. In this author presents the combination of natural language (NL) and machine learning (ML) model to process the dataset which has 1654 aspects in the

training dataset with different category annotations and 845 aspects in the test dataset with different category annotations for analysis.

The performance of software is measured by SVM and logistics regression algorithm. Previously, work was done on the static parameters which reduce the effective learning of tweets according to its label. The nonparametric approach uses number of coefficient parameters those increase the over fitting. In proposed paper, the hyperparameters are used which reduces the adaptive features of learning. The analysis is on the combined performance of parametric (SVM), nonparametric (KNN), and the random forest is used.

## 3 Methodology

The data is collected for the experiment from Twitter and stored in a database for preprocessing. After the preprocessing on the feature label, the processed data is learned by KNN, SVM, and proposed model which is the combination of (KNN, SVM, random forest) these. By these respective models, tweets are predicted and classified, respectively. The below-given section describes the proposed methodology of the model and the techniques used in this work in detail. The pictorial representation of the proposed model is presented in Fig. 1.

**Step 1: Collection of data**
The proposed model used the data collected from the Twitter, regarding government welfare schemes as input. People get easy access to financial and banking services due to Jan Dhan Yojna of government welfare schemes and the main aim of Swachh Bharat Abhiyan is to keep India clean [18]; tweets regarding both the scheme have been used to determine the public review based on aspect-based sentiment analysis. The data that retrieved from the social media is in unstructured form due to intensive information.

**Step 2: Data Storing and fetching**
The retrieved tweets are put in storage as .csv files, and it is fetched using a Python tool PyCharm [19]. Around 3000 tweets are stored for the purpose of training and testing the datasets. Data mining algorithm (SVM, KNN, and hybrid SVM-KNN) is utilized to train and test the fetched tweets.

**Step 3: Preprocessing**
The redundant and noise contents are removed in preprocessing step which makes the data suitable for training process [20].

Data cleaning is performed based on the following steps.

(i) All the uppercase is converted to lowercase.
(ii) Remove all the Internet slangs from the data.
(iii) Removing all the stopping words from the list.

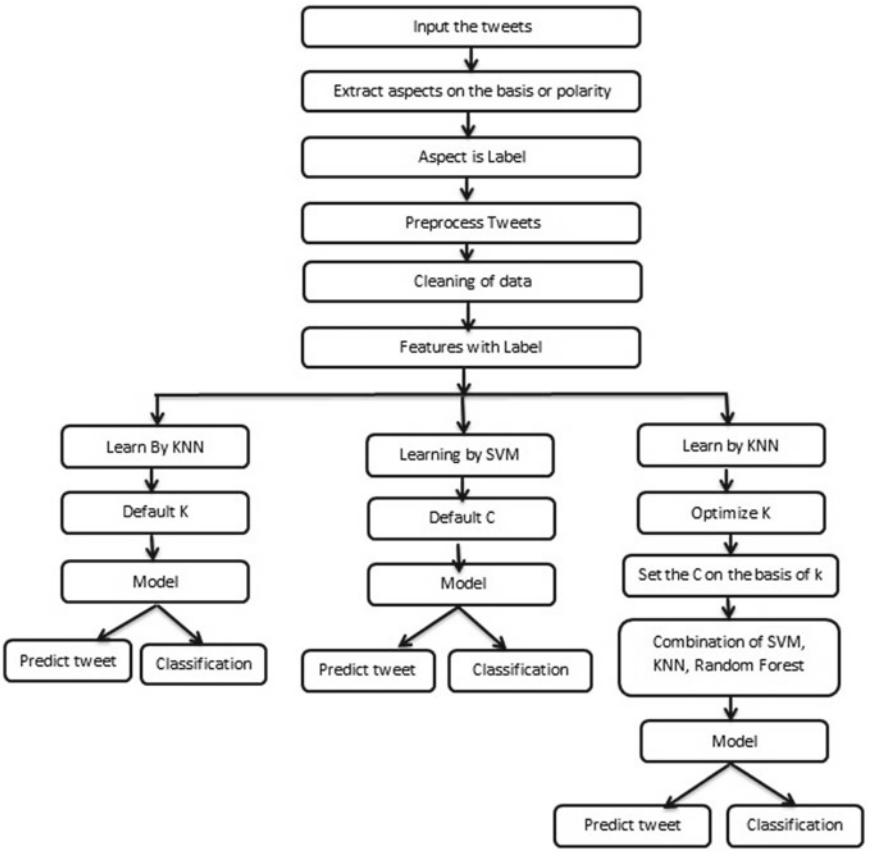



**Fig. 1** Proposed framework process flow

(iv) Eliminating all the additional white spaces.
(v) Compress the duplicate words.
(vi) All the hashtags are removed but the hashtags texts are selected.

**Step 4: Classification using data mining techniques**

Classification is performed using data mining techniques to categorize the data into various aspects such as .

First aspect: increase fund/decrease fund.
Second aspect: improvement in growth/not growth/growth.
Third aspect: goes really fast/works/hard fix.
Fourth aspect: incredible/good/not good.
On the basis of these parameters, the data is trained and tested.

Machine learning models are trained using the training datasets. In the proposed framework, training data is implemented for classification, then testing dataset is prepared that is not before used in the proposed model for training.

**Step 5: Optimize the classification results**
The classification results are need to be checked to confirm whether the learning models training datasets follow the defined rules or not. This process is performed to obtain accurate and error-free results. Python is used to train and test the proposed KNN-SVM random forest model. The tweets nature is predicted as optimal value based on the classification results.

## 4 Results and Discussion

The proposed model classification performance is evaluated and compared with conventional approaches to validate the better performance. The experiment result validation is done by using fivefold and tenfold validation approach. The data is divided into k subsets in cross validation which has equal size and the training process and testing process are repeated for k-times. Every time one group of subset will be used for testing of data and other k-1 subsets of data will be treated as training data. The result analysis is done on SVM, KNN, and the combination of proposed model (SVM, KNN and random forest algorithm). The parameters used in this for analysis process are accuracy, precision, recall, and F-measure.

The classification accuracy of all the three models is depicted in Fig. 2. It is observed that the proposed shows the maximum accuracy in both fold testing process. KNN obtains the least accuracy performance of 49.23%, 51.23%, respectively, for both processes.

The precision analysis for proposed model and conventional SVM and KNN model is depicted in Fig. 3. The proposed algorithm shows the maximum precision and KNN obtains minimum precision 50.23, 51.23 in fivefold and tenfold validation process, respectively (Fig. 4).

The overall performance of all the classifiers is presented in figure. It is observed that the proposed model (KNN-SVM-random forest) shows the optimum, enhanced results with 82% in comparison with the SVM and KNN.

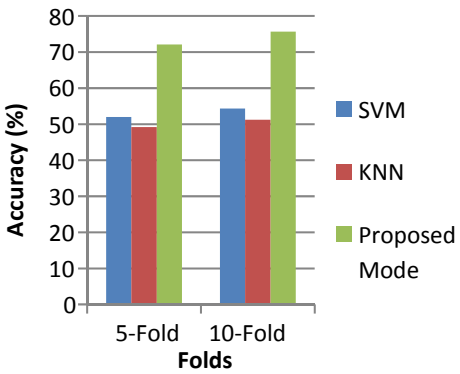



**Fig. 2** Accuracy comparison

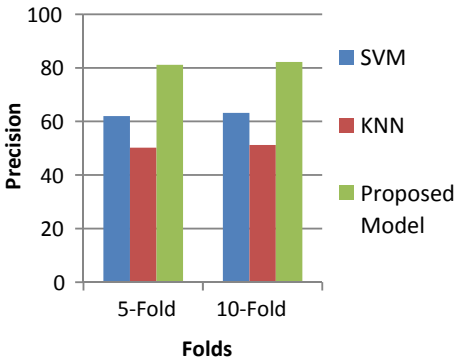



**Fig. 3** Precision analysis

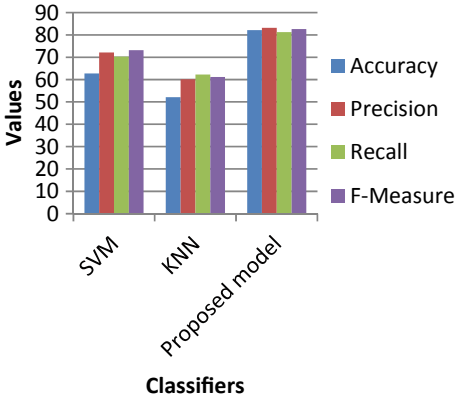



**Fig. 4** Overall performance comparison

## 5 Conclusion

A hybrid classifier is presented in this research work as sentiment analysis model. The proposed hybrid model obtains better results by utilizing part-of-speech (POS) tags and word dependencies for aspect-based sentiment analysis. Efficacy of government welfare schemes using the proposed model is computed 82% under the restricted environment. In future work, improvement in the accuracy will be considered by reducing the feature sparsely by divergence and optimization approaches, improve the classifier by deep learning approaches. Additionally, other tasks that are Aspect Based Sentiment analysis also tested on the applicability on best of learning and concerns with the integration of POS tags, word dependencies, and possibly other NLP tools.

## References


1. Perikos I, Hatzily geroudis I (2017) Aspect based sentiment analysis in social media with classifier ensembles. In: 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS), Wuhan, China, pp 273–278
2. Tiwari SK (2014) To study awareness of a national mission: Swachh Bharat: Swachh Vidyalaya in the middle school student of private and public schools. Paripex-Indian J Res 3(12):23–24
3. Barhate GH, Jagtap VR (2014) Pradhan Mantri Jan Dhan Yojana: national mission on financial inclusion. Indian J Appl Res 4(12):340–342
4. Lin Y, Zhang J, Wang X, Zhou A (2012) An information theoretic approach to sentiment polarity classification. In: Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality, Lyon,France, 2012, pp 35–40
5. Shidaganti G, Hulkund RG, Prakash S (2017) Analysis and exploitation of Twitter data using machine learning techniques. In International proceedings on advances in soft computing, intelligent systems and applications, pp 135–146
6. Rout JK et al (2018) A model for sentiment and emotion analysis of unstructured social media text. Electr Commerce Res 18(1):181–199
7. Mumtaz D, Ahuja B (2017) A lexical and machine learning-based hybrid system for sentiment analysis. In: Studies in computational intelligence. Springer, Singapore (Chap. 11, pp 165–175)
8. Al-Smadi M, Talafha B, Al-Ayyoub M, Yaser J (2018) Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. Int J Mach Learn Cybern 1–13
9. Chen Y (2018) TagNet: toward tag-based sentiment analysis of large social media data. In: 2018 IEEE Pacific visualization symposium (PacificVis), Kobe, Japan, pp 190–194
10. Fouad Mohammed M, Gharib Tarek F, Mashat Abdulfattah S (2018) Efficient Twitter sentiment analysis system with feature selection and classifier ensemble. In: International conference on advanced machine learning technologies and applications, vol 723, pp 516–527
11. Mary Jones T, DivyaSri S, Bavani G (2017) A study on the implications of Pradhan Manthri Jan Dhan Yojana on the growth of indian economy. IRA-Int J Manag Soc Sci 6(3):461–466
12. Khan FH, Bashir S, Qamar U (2014) TOM: Twitter opinion mining framework using hybrid classification scheme. Decis Support Syst 57:245–257
13. García-Pablos A, Cuadros M, Rigau G (2018) W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Syst Appl 91:127–137
14. Pham D-H, Le AX (2018) Learning multiple layers of knowledge representation for aspect based sentiment analysis. Data Knowl Eng 114:26–39
15. Rathan M, Hulipalled VR, Venugopal KR, Patnaik LM (2018) Consumer insight mining: aspect based Twitter opinion mining of mobile phone reviews. Appl Soft Comput 68:765–773
16. Schouten K, van der Wejide O, Frasincar F, Dekker R (2018) Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. IEEE Trans Cybern 48(4):1263–1275
17. Pannala NU, Nawarathna CP, Jayakody JTK, Rupasinghe L, Krishnadeva K (2017) Supervised learning based approach to aspect based sentiment analysis. In: 2016 IEEE international conference on computer and information technology (CIT), Nadi, Fiji, pp 662–666
18. Thakkar P (2015) Swachh Bharat [CLEAN INDIA] Mission—an analytical study. Renew Res J 3(2):168–173
19. Massiris MM, Dennehy BR, Delrieux CA, Thomsen FSL (2017) Python implementation of local intervoxel-texture operators in neuroimaging using Anaconda and 3D Slicer environments. In: 2017 XLIII Latin American Computer Conference (CLEI), Cordoba, Argentina, pp 1–3
20. Dwivedi SK, Rawat B (2016) A review paper on data preprocessing: a critical phase in web usage mining process. In: 2015 international conference on green computing and Internet of Things (ICGCIoT), Noida, India, pp 506–510

# Author Index





A. Pasumpon Pandian et al. (eds.), *Computer Networks, Big Data and IoT*, Lecture Notes on Data Engineering and Communications Technologies 66,
https://doi.org/10.1007/978-981-16-0965-7