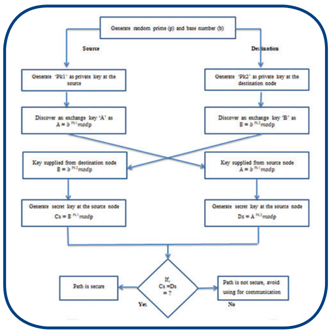
Volume 13, Number 4, 2022

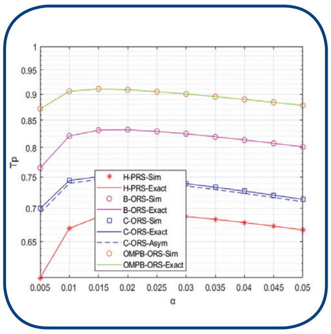
ISSN 1847-6996

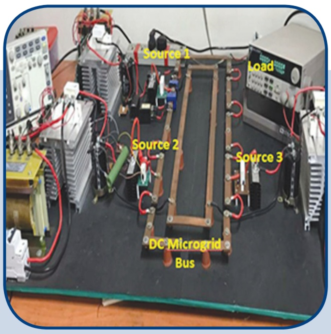
FERIT
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY **OSIJEK**

IJECES International Journal of Electrical and Computer Engineering Systems

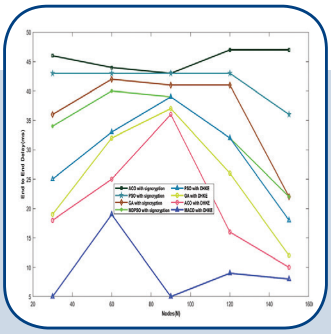
International Journal of Electrical and Computer Engineering Systems


Generate random prime (p) and base number (b)
Source
Destination
Path is secure
Path is not secure, avoid using for communication
Yes
No





Map
Satellite
Sh Abdulla Ave
Manama - Isa Alkabeer Avenue - 4
Road No 720
Road No 722
Manama - Shaikh Isa Avenue - 1
Road No 105



ISSN: 1847-6996

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia

Osijek, Croatia | Volume 13, Number 4, 2022 | Pages 245 - 330

The International Journal of Electrical and Computer Engineering Systems is published with the financial support of the Ministry of Science and Education of the Republic of Croatia

CONTACT

**International Journal of Electrical
and Computer Engineering Systems
(IJECES)**

Faculty of Electrical Engineering, Computer
Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b, 31000 Osijek, Croatia
Phone: +38531224600, Fax: +38531224605
e-mail: ijeces@ferit.hr

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.
Giro account: 2390001 - 1100016777,
Croatian Postal Bank

Proofreader

Ivanka Ferčec
J.J. Strossmayer University of Osijek, Croatia

Editing and technical assistence

Davor Vrandečić
J.J. Strossmayer University of Osijek, Croatia

Stephen Ward
J.J. Strossmayer University of Osijek, Croatia

Dražen Bajer
J.J. Strossmayer University of Osijek, Croatia

Journal is referred in:

- Scopus
- Web of Science Core Collection (Emerging Sources Citation Index - ESCI)
- Google Scholar
- CiteFactor
- Genamics
- Hrčak
- Ulrichweb
- Reaxys
- Embase
- Engineering Village

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003
Published: quarterly
Circulation: 300

IJECES online

https://ijeces.ferit.hr

TABLE OF CONTENTS

MACO-DHKE Based Secure Data Transmission in MANETs...245

Original Scientific Paper

Sandeep Dhende | Suresh Shirbahadurkar

Distribution and Allocation of Network Resources
Based on Predictive Analyses of Time-Series Telecommunications Data 253

Original Scientific Paper

Višnja Križanović | Jelena Vlaović | Drago Žagar | Snježana Rimac-Drlje

An Optimized Distributed Routing Protocol
for Energy Management Systems Based on Wireless Sensor Networks
in Intelligent and Smart Structures...269

Original Scientific Paper

Ahmed Hammad | M.A. Mohamed | Heba M. Abdel-Atty

Development and Hardware Implementation
of IoT-Based Patrol Robot for Remote Gas Leak Inspection..279

Original Scientific Paper

Tasneem Yousif | Wael El-Medany

A Trust-Based Recommender System
for Personalized Restaurants Recommendation 293

Original Scientific Paper

Qusai Shambour | Mosleh M. Abualhaj | Ahmad Adel Abu-Shareha

Bolstering user authentication:
a kernel-based fuzzy-clustering model for typing dynamics 301

Original Scientific Paper

Anthony Metumaraibe Nwohiri | Ufuoma Cyril Ogude | Hai Vinh Nguyen
Edilberto Moreno Sanchez

A Novel Method for Real Time Protection
of DC Microgrid Using Cumulative Summation and Wavelet Transform 311

Original Scientific Paper

Dipti D Patil | Bindu S | Sushil Thale

Classification of Healthcare Service Reviews
with Sentiment Analysis to Refine User Satisfaction 323

Review Paper

Khai Herng Leong | Dahlila Putri Dahnil

About this Journal

IJECES Copyright Transfer Form

MACO-DHKE Based Secure Data Transmission in MANETs

Original Scientific Paper

Sandeep Dhende
Research Scholar, Department of Electronics and Telecommunication Engineering,
ZES's Zeal College of Engineering and Research, Pune, SPPU, Maharashtra, India
sandeepldhende@gmail.com

Suresh Shirbahadurkar
Professor, Department of Electronics and Telecommunication Engineering,
ZES's Zeal College of Engineering and Research, Pune, SPPU, Maharashtra, India
shirsd@gmail.com

Abstract *– Mobile ad hoc networks (MANETs) are the self-sufficient nodes with their unique characteristics, such as open wireless mediums and self-motivated topology, which suffers from security weaknesses. Due to the complexity of the MANET security architecture, it is difficult to detect and prevent security issues in the wired networks. Hence, this paper proposes a secure and QoS-aware data transmission in the MANET that improves the efficiency of the transmission. The optimal route, which is the shortest possible path, is chosen using the modified ant colony optimization algorithm (MACO algorithm) and the secure transmission in MANET is ensured using the Diffie-Hellman key-exchange algorithm (DHKE strategy). The key exchange method is used to improve the security of the transmission of the MANET. This paper shows a high transmission rate on a secure path with a higher PDR and throughput.*

Keywords*: MANET, Security, MACO, DHKE, Routing*

1. INTRODUCTION

The MANET network is a distributed multi-hop network composed of mobile nodes that can be used for various applications [1] [2]. These nodes can interact with each other in a multi-hop fashion. Each of the nodes in a network is a part of the same network, which means that they play an equal role in the transmission of data. The mobile devices in a network are known as routers, which are used to control and route the data packets. The routing protocol serves as a way for a network to establish contact with its source and destination. It also determines the best path for a packet to reach its intended destination. The MANET communication is investigated using several network parameters, such as packet delivery ratio (PDR), delay, distance, and so on, which are enclosed within the Quality of service (QoS) of the network [3]. The optimal route for availing quality-of-service is chosen using the following factors: MACO, HRM, and energy efficiency [2].

Outstanding routing techniques were developed with the goal of eliminating wasteful data transfer losses [20] [22]. Routing methods reduce routing communication to a bare minimum to ensure efficiency, but the resulting minimal data creates a barrier for a Routing Attack Detection System (RADS) [21]. The main issue with course disclosure is that it can result in loss of data. Fortunately, there are various algorithms that can help minimize this issue. Some of these include the Genetic algorithm, simulated annealing, and particle swarm optimization [2]. Secure routing techniques have been developed to protect MANETs from attacks, but these protocols do not protect additional data [23]. Because wireless sensor networks may be used in hostile environments such as battlefields [24] security in data communication is an important problem to consider when developing them. The Bee Ad Hoc fuzzy logic (FBeeAdHoc) framework has been used to provide a security layer for routing protocols in MANETs. In MANETS, the Homomorphic Encryption (HE) approach is used to accomplish multi-level security, which means that the data will be subjected to a cryptographic hash function and an encryption algorithm will be sent to the end, where the data will be decrypted and the data's dependability will be checked. The wormhole attack is one of the most serious security threats that can significantly disrupt network connectivity. When a node breaks the security standards and so becomes vulnerable to attack, this is referred to as malicious behavior [25].

The paper proposes a routing protocol that can increase distribution ratios and reduce end-to-end delay. It also uses the Diffie-Hellman algorithm to improve security of path. The paper is divided into four sections: Section 2, Section 3, Section 4, and Section 5. The topics covered by these sections include: (1) an overview of the concepts of routing and data protection (2) a comparison of various secure routing and protection protocols (3) a description of the characteristics of the proposed method.

2. LITERATURE REVIEW

In this section, the review of the existing literature is presented with the upcoming challenges of the research. An approach was proposed in 2019 to find the best CH for a stable MANET security level. The proposed model combines the number of attackers and the approximate distance and TV consumption of each certificate holder to ensure that the data is available and stable [1]. In 2019, Mariappan Rajashanthi and K. Vathi proposed a secure multipath routing scheme that is energy-efficient and has an encryption technique [4]. In 2018, the authors of this paper proposed various security-related regulations for data collection. They then reviewed the various methods and techniques related to MANET detection [5]. In 2017, the two authors updated the LEACH and AOMDV routing strategies. They provided a uniform approach for multipath routing and cluster generation [6]. In 2017, mobile ad hoc networks (MANETs) were introduced to the concept of security through the use of pre-existing routing protocols. The goal is to gain fast and secure communication while protecting the integrity and authenticity of the network [7].

The goal of a network is to receive and send messages from two users. To achieve this, the algorithm used to set the routing path was optimized. It allowed for better packet delivery and lower EC [8]. In 2018 [9], Mostafaei recommended a disseminated learning machine based calculation to work on the organization's exhibition with a few obliged QoS boundaries. It took a couple of QoS directing limitations into the record in way choice, like start to finish steady quality and deferral. As far as start to finish postponement and energy-viability, the outcomes showed that the estimation performed better compared to the present status of the craftsmanship brutal computations.

To decide the best area of the gathering particles, the fundamental multitude streamlining was refreshed. On account of a steady organization structure, one molecule is relied upon to decide the G-best position, and the leftover particles can search for additional spaces to confirm that the best position is G-best, not the flow one. In this article, Modified Ant Colony Optimization (MACO) is used to tackle the inadequacies of current renditions of MANET. Most of past research has focused on either energy productivity or unwavering quality; in any case, in this article, both energy-effective bunching and dependability are joined in a solitary MANET model.

3. PROPOSED SECURE TRANSMISSION IN MANET USING THE MACO ALGORITHM AND DHKE STRATEGY:

The method of sending information from a source to an objective without the need of a wired media is known as remote correspondence. A portion of the WSN's and MANET's elements are practically indistinguishable. In the modern days, MANET plays a significant role in rendering the network services equipped within the handheld devices. Hence, there is a need to utilize the routing protocol for ensuring the easy access to the network services, where the optimal route is decided for reaching the services available through the hand-held devices. In this research, MACO algorithm is proposed for selecting the optimal communication path in MANET. Moreover, rendering security for the data provider and user is very significant, which is ensured using the DHKE strategy. The MACO algorithm and DHKE strategy are utilized in the MANET, which further promotes the QoS of the network.

3.1 MANET COMMUNICATION

The network is equipped with numerous sensor nodes, which are engaged in the data transmission in the network as shown in figure 2. Following table 1 shows the network parameters employed for simulating the MANET in NS2. The source node generates the request message for initiating the communication with the destination node, and the communication is preceded only if the security keys of both the source and destination nodes matches with each other. In this context, optimal route selection and selection of the secure path is the major focus.

3.2 PROPOSED MODIFIED ANT COLONY OPTIMIZATION (MACO) FOR OPTIMAL PATH SELECTION IN MANET:

In the MANET, ensuring the throughput rates is important to meet the client demands with an effective QoS. Due to different plan hardships and imperative satisfaction, conventional conventions fail to address the user challenges. Hence, upgrading throughput turns into a basic issue to fulfill client needs and application support. Therefore, throughput is the significant factor for rendering the required QoS for any kind of MANET applications and in this research, MACO streamlining technique considers throughput as one of the factor in selecting the optimal routing path for MANET communication.

3.2.1 Solution representation:

In a optimization algorithm, the solution representation signifies the solution declared by the algorithm. In this research, solution is the routing path with the source node as the initiating node or the data sender S and destination node D as the terminating node or the receiver, with the intermediating nodes $(I_1, I_2, ... I_n) such that (n<m)$ being the communicating nodes between the source and destination nodes as shown in figure 1.

S	I_1	I_2	………………	D

Fig. 1. Solution representation

where, *m* is the total nodes in the MANET with *n* being the intermediate nodes in the communicating nodes.

3.2.2 Fitness measure:

The optimal solution, which is the optimal routing path, is decided by MACO using the fitness measures, such as throughput, PDR, routing overhead, and delay. The solution is selected as optimal when the throughput and PDR is high with the minimal overhead and transmission delay.

3.2.3 MACO description:

The MACO algorithm is the modified version of the ACO algorithm, which aims at the selection of the optimal route between the source and the destination nodes. The optimal route is the solution of MACO as per the figure 1. Technically, the solution or the route refers to the ants in the MACO and initially, the proposed MACO establishes the random solutions at the initial iteration, which is accompanied with the generation of all the possible routes between the source and the destination nodes, from which the optimal route satisfying the fitness measure is selected for communication.

The MACO builds the connection's packet transmission rate, bringing about a reasonable course choice arrangement. Forward ant is begun by the source hub at arbitrary to visit the entirety of the open hubs in the course [15]. During their crossing, the ants leave a little amount of pheromone on the visited joins. At the point when the ants show-up at their objective, the ants update the pheromone of all hubs visited all through the crossing. A hub's throughput is treated as a pheromone for this situation. The throughput work is utilized to refresh a hub's pheromone [16] [17].

Equation 1 is used to calculate $f(t)$.

$$f(t) = \max \sum_{i=1}^{k} \frac{p(i)}{t(i)} \qquad (1)$$

Where k denotes the packet transmission limit, $p(i)$ is the number of packets successfully transferred, and $t(i)$ denotes the packet transmission time.

An ant (A) is a collection of routes that link all nodes. MACO's fitness function shown in equation 2, also known as the objective function, is shown as follows:

$$\text{fitness of ant} = \sum_{i=1}^{n-1} d(i,j) \quad \forall\, j = n \;\Lambda\; j = i+1 \qquad (2)$$

The pheromone is updated in a cyclic way during the course of each traversal of a link l. Equation 3 is used to calculate the likelihood of an ant 'm' visiting node 'j' from node i.

$$\rho_{ij}^{d}(t) = \frac{[\tau_{ij}(t)]^{\alpha}.[\mu_{ij}(t)]^{\beta}.[e_j(t)]^{\gamma}}{\sum_{j\in N}[\tau_{ij}(t)]^{\alpha}.[\mu_{ij}(t)]^{\beta}.[e_j(t)]^{\gamma}} \qquad (3)$$

The pheromone concentration in link ij is τ_{ij}, e_j is the energy of the node, control parameters are α,β and γ, and the throughput heuristic value μ_{ij} is $f(t)$.

Equation 4 is used to calculate the pheromone concentration as it decreases over time.

$$\tau_{ij} = (1-\rho) * \tau_{ij} + \sum_{n=1}^{m} \Delta_{ij}{}^{n}$$

Where, Δ_{ij}^{n} is the change in pheromone amount in the link ij, updated by the mth ant, and (1-ρ) is a decreasing pheromone constant. The following generation of ants migrates to their goal via increasing pheromone concentration nodes.

This cycle is proceeded until the state of stagnation is satisfied. The street that arises after a time of balance is viewed as the best way for correspondence. This methodology is done for every information transmission. This progression flags the beginning of the organization's transmission interaction. The figure 2 portrays the most limited way that is discovered utilizing AODV considering every one of the elements of MACO improvement calculation.

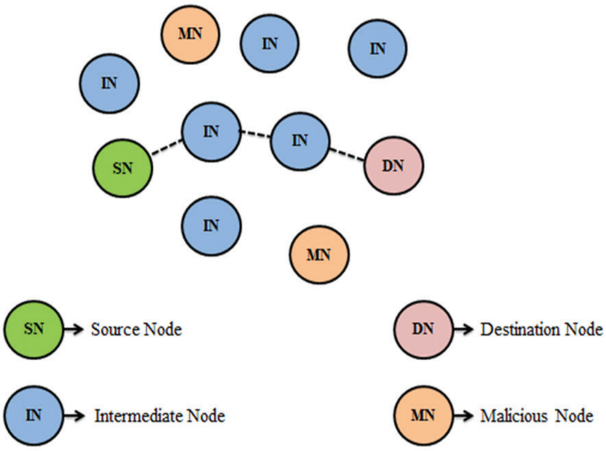


Fig. 2. MANET network

3.3 DHKE STRATEGY FOR THE SECURE COMMUNICATION IN MANETS

Using a finite number of nodes, a MANET is simulated in NS2, where the nodes communicate the data between the nodes only when the routing path offers better QoS with guaranteed security. Thus, security is ensured through DHKE strategy, which ensures the path is secure. Initially, find the source and destination nodes in the MANET for transferring the data packets between the nodes. The AODV routing protocol is used with MACO to identify the shortest path between these nodes. The shortest route from source to destination is chosen using MACO, where the security is guaranteed using DHKE [19].

First and foremost, a conduit from source to destination is constructed followed with the data transfer. Nodes at the source and destination locations for data transfer create two random numbers, p (prime number) and b (base number) in a DHKE-based strategy. The source node generates the private key Pk1 and the destination node generates the private key Pk2. Using the following formulas, two values A and B are calculated on the source and destination ends, respectively.

$$A = \mathrm{bPk1modp} \tag{5}$$

$$B = \mathrm{bPk2modp} \tag{6}$$

The values of A and B are exchanged across the nodes in order to calculate the secret key values at both ends. The formulae for computing the value of the secret key at the source and destination nodes are as follows:

$$\mathrm{Cs} = \mathrm{BPk1modp} \tag{7}$$

$$\mathrm{Ds} = \mathrm{APk2modp} \tag{8}$$

These secret key values are compared with each other for enabling the secure data transfer. Upon the mismatch in the secret keys between the source and destination nodes, the packets drop intimating the presence of the malicious nodes in the network thereby, blocking further communication. The DHKE strategy for secure path selection is depicted in figure 3 as a flowchart.

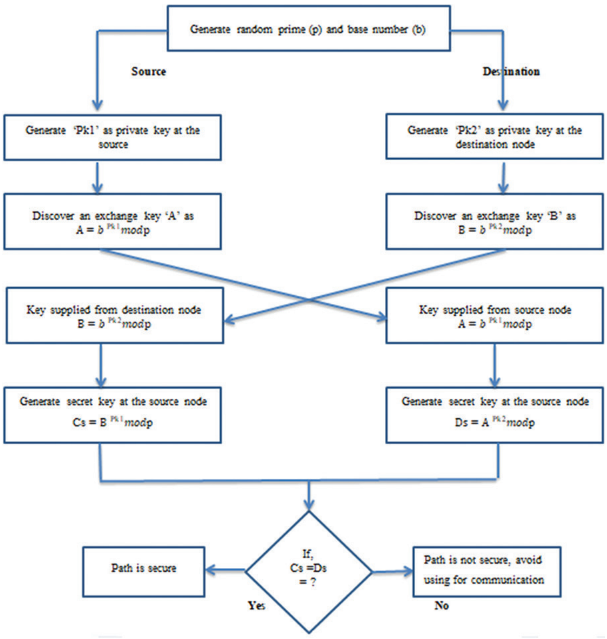

Generate random prime (p) and base number (b)
Source
Des ination
Generate 'Pk1' as private key at the source
Generate 'Pk2' as private key at the destination node
Discover an exchange key 'A' as $A = b^{Pk1} mod p$
Discover an exchange key 'B' as $B = b^{Pk2} mod p$
Key supplied from destination node $B = b^{Pk2} mod p$
Key supplied from source node $A = b^{Pk1} mod p$
Generate secret key at the source node $Cs = B^{Pk1} mod p$
Generate secret key at the source node $Ds = A^{Pk2} mod p$
If, Cs =Ds = ?
Path is secure
Path is not secure, avoid using for communication
Yes
No

Figure 3. Diffie-Hellman approach for checking whether path is secure or not

4. RESULT ANALYSIS

In this section, the achievements of the MACO with DHKE strategy are portrayed in order to enumerate the effectiveness compared with the existing state-of-art methods.

4.1 SIMULATION ENVIRONMENT:

The simulation is established in NS2 environment with the network settings shown in table 1. In the MANET network, a maximal of 150 nodes are distributed in the simulation area of coverage 1500 m × 1500 m.

Table 1. Parameters for simulation

Parameters	Network Settings
Number of Nodes	30, 60, 90, 120 and 150
Area Size	1500 m × 1500 m
Transmission Range	250 m
Data Types	CBR
Packet Size	512Bytes
Antenna	Omni directional
Type of Queue	Drop Tail
Routing protocol	AODV

4.2 PERFORMANCE METRICS:

The effectiveness of the routing protocol, MACO with DHKE strategy is revealed through the analysis based on the metrics, such as packet delivery ratio (PDR), throughput, routing overhead, and delay.

4.3 COMPARATIVE ANALYSIS:

The methods employed for the comparative analysis include: ACO, genetic algorithm (GA), particle swarm optimization (PSO), and MDPSO [1]. The difference in the packet delivery ratio (PDR) with respect to the quantity of hubs is displayed in Figure 4. The proposed MACO with DHKE acquires better PDR, throughput, delay and overhead when compared with the existing methods, like ACO with signcryption, PSO with signcryption, GA with signcryption, MDPSO with signcryption, PSO with DHKE, GA with DHKE, and ACO with DHKE. Table 2 shows the acquired throughput for the methods. The PDR, throughput, overhead and delay analysis is enumerated in figures (4) – (7) and tables (2) – (5).

The PDR analysis is performed with respect to the number of the nodes (in table 2), where it is highlighted that the PDR percentage shows slight improvement with the increasing number of nodes. Though the PDR decreases with the increasing number of nodes due to link failure, the application of the secure path selection method boosts the PDR through enhancing the link lifetime of the network. The proposed MACO with DHKE acquired the PDR of 97% when 150 nodes are communicating in the network, which is the best ever acquired PDR percenatge, which is mainly due to the development of the secure path selection mechanism.

The throughput analysis of the methods based on the total nodes is demonstrated in the figure 5. The throughput of the methods are affected when the transmission overhead prevails in the network due to the total number of users. When the total nodes is 150, the throughput acquired by the proposed MACO with DHKE is 4367 kbps, which is better when compared with the existing methods, justifying the effectiveness of the proposed method.

Table 2. PDR analysis (in %) (Higher PDR is better)

Number of nodes	ACO with signcryption	PSO with signcryption	GA with signcryption	MDPSO with signcryption	PSO with DHKE	GA with DHKE	ACO with DHKE	**proposed MACO with DHKE**
30	85	89	91	91	92	93	94	**94**
60	85	85	86	86	91	92	92	**94**
90	85	86	86	90	94	94	95	**96**
120	85	86	89	93	95	95	95	**95**
150	86	88	88	92	94	95	96	**97**

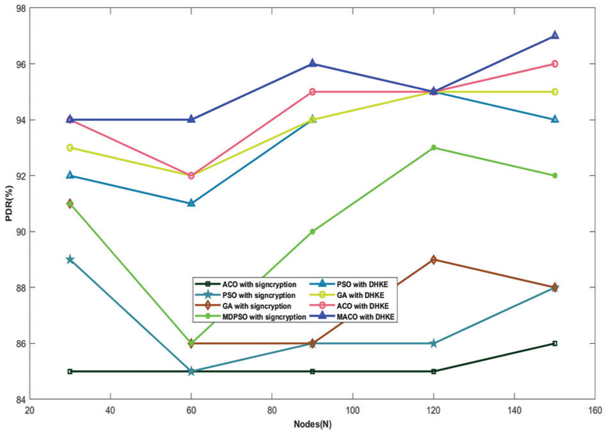


Fig. 4. Analysis based on PDR (Higher PDR is better)

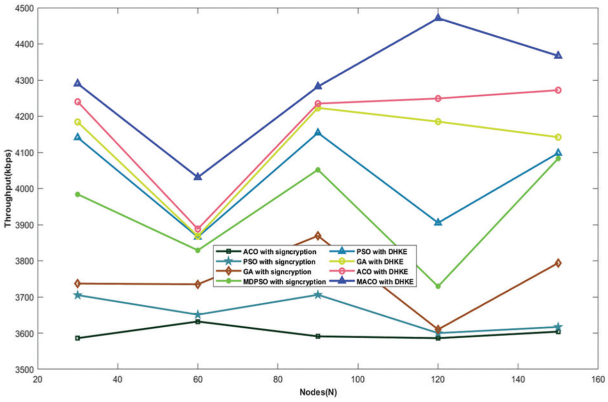


Fig. 5. Throughput analysis (Higher throughput is better)

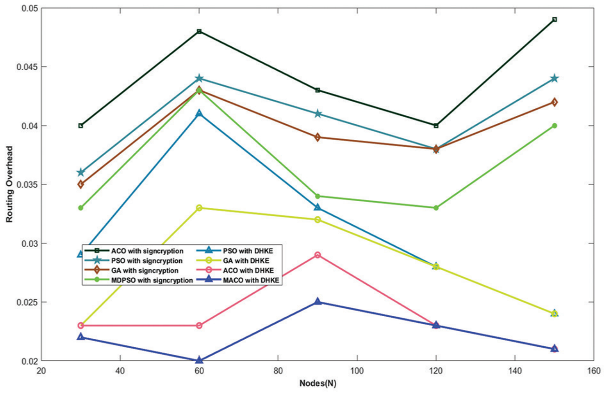


Fig. 6. Overhead analysis (Minimal overhead is better)

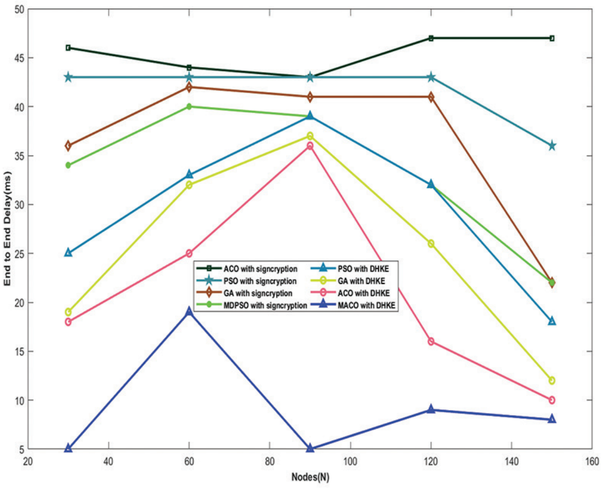


Fig. 7. Delay analysis (Minimal delay is better)

Similarly, the overhead analysis is performed, which focuses on the analysis of the computational complexity of the network when the node communication increases. When the total nodes is 30, the overhead is minimal, while the overhead incrases with the increase in the total simulated nodes in the network. However, when compared with the existing methods, the computational overhead of the proposed model is minimal, which insists that the proposed method schedules the communication through the optimal path.

Likewise, the delay analysis in the figure 7 insists that the effective performance of the network is based on the minimal delay of data communication between the nodes. For instance, as mentioned in the table 5, when the network is simulated with 150 nodes, the network communication delay is found to be around 8ms while using the proposed MACO with DHKE, which shows the siginificance of the proposed method in exhibiting the effective performance.

In short, the method that exhibts the minimal delay, minimal communication overhead, higher throuhgput and higher PDR are the best method. The proposed MACO with DHKE outperforms the existing methods with the minimal delay of 5ms with 30 and 90 nodes, and the minimal overhead of 0.02. On the other hand, the maximal PDR and throughput acquired by the proposed MACO with DHKE is 97% (with 150 nodes) and 4471 kbps (with 120 nodes), respectively.

Table 3. Throughput analysis (in kbps) (Higher throughput is better)

Number of nodes	ACO with signcryption	PSO with signcryption	GA with signcryption	MDPSO with signcryption	PSO with DHKE	GA with DHKE	ACO with DHKE	**proposed MACO with DHKE**
30	3586	3705	3737	3984	4141	4184	4240	**4290**
60	3632	3651	3735	3829	3866	3868	3888	**4031**
90	3591	3706	3869	4052	4154	4223	4235	**4282**
120	3586	3600	3610	3729	3905	4185	4249	**4471**
150	3604	3617	3794	4083	4098	4142	4272	**4367**

Table 4. Overhead analysis (Minimal overhead is better)

Number of nodes	ACO with signcryption	PSO with signcryption	GA with signcryption	MDPSO with signcryption	PSO with DHKE,	GA with DHKE	ACO with DHKE	**proposed MACO with DHKE**
30	0.04	0.036	0.035	0.033	0.029	0.023	0.023	**0.022**
60	0.048	0.044	0.043	0.043	0.041	0.033	0.023	**0.02**
90	0.043	0.041	0.039	0.034	0.033	0.032	0.029	**0.025**
120	0.04	0.038	0.038	0.033	0.028	0.028	0.023	**0.023**
150	0.049	0.044	0.042	0.04	0.024	0.024	0.021	**0.021**

Table 5. Delay analysis (in ms) (Minimal delay is better)

Number of nodes	ACO with signcryption	PSO with signcryption	GA with signcryption	MDPSO with signcryption	PSO with DHKE,	GA with DHKE	ACO with DHKE	**proposed MACO with DHKE**
30	46	43	36	34	25	19	18	**5**
60	44	43	42	40	33	32	25	**19**
90	43	43	41	39	39	37	36	**5**
120	47	43	41	32	32	26	16	**9**
150	47	36	22	22	18	12	10	**8**

5. CONCLUSION:

With the aid of an MACO with DHKE approach, the researchers suggested a unique model for a secure transmission in the MANET. The MACO-based MANET routing was utilized to find the best and shortest path in the network. For data security in the MANET, the DHKE strategy is used, which determines whether the path between the source and destination nodes is safe. Furthermore, the suggested model combined the number of attackers with the performance evaluation procedure to examine the number of attackers. The performance of the proposed MACO with DHKE is enumerated based on the performance measures, such as delay, throughput, PDR, and overhead. The proposed method exhibts the minimal delay, minimal communication overhead, higher throuhgput and higher PDR of 5ms with 30 and 90 nodes, and the minimal overhead of 0.02, maximal PDR and throughput of 97% (with 150 nodes) and 4471 kbps (with 120 nodes), respectively. Furthermore, the suggested MACO with DHKE method achieved an improved result. To provide security in the MANET, backup routing in ad hoc networks (AODV) with a DHKE approach might be studied for detection of malicious node in future study. In addition, by combining bio-inspired and security algorithms, the performance of MANETs with security may be increased.

6. REFERENCES:

[1] M. Elhoseny, K. Shankar, "Reliable Data Transmission Model for Mobile Ad Hoc Network Using Signcryption Technique", IEEE Transactions on Reliability, Vol. 69, No. 3, 2020, pp. 1077-1086.

[2] L. Harn, C. F. Hsu, O. Ruan, M. Y. Zhang, “The novel design of secure end-to-end routing protocol in wireless sensor networks”, IEEE Sensors Journal, Vol. 16, No. 6, 2016, pp. 1779–1785.

[3] K. Vijayan, A. Raaza, “A novel cluster arrangement energy-efficient routing protocol for wireless sensor networks”, Indian J. Sci. Technol., Vol. 9, No. 2, 2016, pp. 1–9.

[4] M. Rajashanthi, K. A. Valarmathi, “Secure Trusted Multipath Routing and Optimal Fuzzy Logic for Enhancing QoS in MANETs“, Wireless Personal Communications, Vol. 112, 2020, pp. 75–90.

[5] G. Liu, Z. Yan, W. Pedrycz, “Data collection for attack detection and security measurement in mobile ad hoc networks: A survey”, Journal of Network and Computer Applications, Vol. 105, 2018, pp. 105–122.

[6] B. Rana, D. Rana, “Energy efficient load balancing with clustering approach in MANET”, Proceedings of the International Conference on Energy, Communication, Data Analytics and Soft Computing, Chennai, India, 1-2 August 2017, pp. 2019–2024.

[7] D. Hurley-Smith, J. Wetherall, A. Adekunle, "SUPERMAN: Security Using Pre-Existing Routing for Mobile Ad hoc Networks", IEEE Transactions on Mobile Computing, Vol. 16, No. 10, 2017, pp. 2927-2940.

[8] Y. J. Oh, K. W. Lee, “Energy-efficient and reliable routing protocol for dynamic-property-based clustering mobile ad hoc networks”, Int. J. Distrib. Sens. Netw., Vol. 13, No. 1, 2017.

[9] H. Mostafaei, “Energy-efficient algorithm for reliable routing of wireless sensor networks”, IEEE Transactions on Industrial Electronics, Vol. 66, No. 7, 2019, pp. 5567–5575.

[10] M. Elhoseny, K. Shankar, S. K. Lakshmanaprabu, A. Maseleno, N. Arunkumar, “Hybrid optimization with cryptography encryption for medical image security in Internet of Things”, Neural Computing Applications, 2018, pp. 1–15.

[11] K. Shankar, M. Elhoseny, E. D. Chelvi, S. K. Lakshmanaprabu, W. Wu, “An efficient optimal key based chaos function for medical image security”, IEEE Access, Vol. 6, 2018, pp. 77145–77154.

[12] D. Gupta, A. Khanna, K. Shankar, V. Furtado, J. J. Rodrigues, “Efficient artificial fish swarm based clustering approach on mobility aware energy-efficient for MANET”, Transactions on Emerging Telecommunications Technologies, 2018, Art. No. e3524.

[13] Y. Yang, “Broadcast encryption based non-interactive key distribution in MANETs”, Journal of Computer and System Sciences, Vol. 80, No. 3, 2014, pp. 533–545.

[14] S. B. Prabaharan, R. Ponnusamy, “ Enhanced Longevity of MANETs using ACO based Balanced Network Monitoring and Routing Model (BNMR)”, Advances in Wireless and Mobile Communications, Vol. 10, No. 5, 2017, pp. 1035-1049.

[15] C. Ratanavilisagul, “Modified Ant Colony Optimization with Pheromone Mutation for Travelling Salesman Problem”, Proceedings of the 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Phuket, Thailand, 27-30 June 2017.

[16] S. Kaur, R. Mahajan, “Hybrid meta-heuristic optimization based energy efficient protocol for wireless sensor networks”, Egyptian Informatics Journal, Vol. 19, No. 3, 2018, pp. 145-150.

[17] Y. Gao, J. Wang, W. Wu, A. K. Sangaiah, S.-J. Lim, “A Hybrid Method for Mobile Agent Moving Trajectory Scheduling using ACO and PSO in WSNs”, Sensors, Vol. 19, No. 3, 2019, p. 575.

[18] M. Alotaibi, “Security to wireless sensor networks against malicious attacks using Hamming residue method”, EURASIP Journal on Wireless Communications and Networking, Vol. 8, 2019.

[19] O. Singh, J. Singh, R. Singh, “DHHP: A Hybrid Technique for Protecting Mobile Adhoc Networks from Selective Packet Drop Attack”, International Journal of Computational Intelligence Research, Vol. 13, No. 7, 2017, pp. 1743-1763.

[20] M. Rajashanthi, K. Valarmathi, “A Secure Trusted Multipath Routing and Optimal Fuzzy Logic for Enhancing QoS in MANETs”, Wireless Personal Communications, Vol. 112, 2020, pp. 75–90.

[21] G. Liu, Z. Yan, W. Pedrycz, “Data collection for attack detection and security measurement in mobile ad hoc networks: A survey”, Journal of Network and Computer Applications, Vol. 105, 2018, pp. 105–122.

[22] R. Anita, "Joint cost and secured node disjoint energy efficient multipath routing in mobile ad hoc network", Wireless Networks, Vol. 23, No. 7, 2017, pp. 2307–2316.

[23] D. Hurley-Smith, J. Wetherall, A. Adekunle, "SUPERMAN: Security Using Pre-Existing Routing for Mobile Ad hoc Networks," IEEE Transactions on Mobile Computing, Vol. 16, No.10, 2017, pp. 2927-2940.

[24] K. Vijayan, A. Raaza, "A novel cluster arrangement energy-efficient routing protocol for wireless sensor networks", Indian Journal of Science and Technology, Vol. 9, No. 2, 2016, pp. 1–9.

[25] M. Rafsanjani, H. Fatemidokht, "FBeeAdHoc: A secure routing protocol for BeeAdHoc based on fuzzy logic in MANETs", AEU-International Journal of Electronics and Communications, Vol. 69, No. 11, 2015, pp. 1613–1621.

Distribution and Allocation of Network Resources Based on Predictive Analyses of Time-Series Telecommunications Data

Original Scientific Paper

Višnja Križanović
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Sceince
and Information Technology Osijek
KnezaTrpimira 2b, Osijek, Croatia
visnja.krizanovic@feirt.hr

Drago Žagar
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Sceince
and Information Technology Osijek
KnezaTrpimira 2b, Osijek, Croatia
drago.zagar@feirt.hr

Jelena Vlaović
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Sceince
and Information Technology Osijek
KnezaTrpimira 2b, Osijek, Croatia
jelena.vlaovic@feirt.hr

Snježana Rimac-Drlje
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Sceince
and Information Technology Osijek
KnezaTrpimira 2b, Osijek, Croatia
snjezana.rimac@feirt.hr

Abstract *– With the fast development of different communication technologies, applications, and services, the adoption of advanced sensory and computing solutions, such as the various Internet of Things (IoT) and mobile computing solutions, is continuously growing. The massive adoption of mobile computing and IoT sensory devices encouraged the continuous growth of generated network traffic. Therefore, the selection of adequate solutions for efficient data processing became necessary. Despite numerous advantages arising from effective data processing, operators and enterprises working within the ICT domain have only limited amounts of available networking resources to store, process, and use valuable information extracted from large quantities of gathered data. In this paper, an optimal planning process and prediction of usage of network resources is examined. It takes into consideration the results of predictive modeling processes based on available sets of time series telecommunications data. The given forecasts enable effective selection of network architectures, as well as the distribution and allocation of network resources considering the cloud, edge, and fog networking concepts.*

Keywords*: optimal network resources allocation; edge-fog-cloud management for networking load distribution; telecommunications time-series data analyses; predictive analyses*

1. INTRODUCTION

The significant advances in the development of information and communication technologies (ICT) have led to the massive adoption of various mobile computing and sensory devices, as well as an exponential growth of generated wireless network traffic [1]-[6]. A constant evolution and adoption of advanced technological solutions, such as the Internet of Things (IoT) solutions, expands opportunities for implementation of the Internet of Everything (IoE) ecosystems. These intelligent ecosystems are comprised of a large number of distributed heterogeneous devices that consistently generate vast amounts of data.

According to some estimates, within a few years, there will be several hundred billion connected IoT devices affecting every aspect of life, ranging from personal, public or industrial smart environments [1]. In this context, sensors have a crucial role in enabling the collection of various types of acquired data [7]. For handling massive amounts of data originating from numerous sources, increasing attention is given to the efficient transfer of network traffic and data processing techniques [1]. Overall growth in the volume of network traffic and the development of the next generation ICT ecosystems encourage telecom operators and enterprises in the ICT domain to adapt their existing networking approaches, [4] and [6]. Adjusting network settings can achieve efficient distribution and allocation of resources for caching, processing, and computing. Due to existing requirements related to the performance of novel ICT systems, optimal selection of adequate network architectures became a very challenging task. The processes of appropriate network resource distribution and allocation significantly impact data processing speed. Therefore, these processes can be additionally supported by the implementation of algorithms for adaptive resource management, taking into consideration data usage patterns, as well as adequate techniques for traffic offloading [1]. The processes of prediction-based planning, distribution, and allocation of network resources present significant challenges. The selection of adequate predictive models can effectively contribute to the optimization of planning processes, as can also be seen based on the results of analyses presented in this paper. Telecommunications data analytics usually takes into consideration time-series data reflecting adoption or deployment rates of particular ICT solutions. The gathered results point to specific requirements that are necessary for optimization. If data representing ICT solutions' adoption and deployment patterns are taken into consideration to derive useful knowledge, suitable prediction models must be used. This aspect is particularly important because a vast number of telecom operators and enterprises in the ICT domain search for the best solutions for the distribution and allocation of their limited caching and computing resources.

In this paper, the analytics of gathered available telecommunications data sets are conducted based on the results created using several predictive models. The smaller sets of time series data are taken into consideration. In Section 2, an overview of current trends in adoption and deployment of relevant telecommunications solutions is presented, as a starting point for further prediction-based forecasts and planning processes. In Section 3, the differences between centralized and decentralized network architectures, and between edge and fog computing concepts are presented. In Section 4, an overview of several common and some additional predictive models are presented, and the usage of gathered results in processes of network resources allocation and distribution is accentuated. In Section 5, the defined models are applied to collected data sets, and the obtained results are carefully analyzed. The optimal approaches in making decisions related to network resources distribution and allocation are indicated based on predictive analyses results, and the most important conclusions are highlighted.

2. ADOPTION TRENDS OF TELECOMMUNICATIONS SOLUTIONS

The telecom operators and enterprises working within the ICT domain have high expectations for scale and scope arising from advanced ICT solutions offerings [1]. The additional information extracted from a large amount of collected network traffic presents added value that encourages a significant incentive for advanced ICT services development and implementation processes. These expectations induce new research challenges related to available network settings. Higher levels of availability and quality of ICT solutions induce additional growth of their adoption, which is closely correlated with total achieved gains. The majority of businesses based upon usage of advanced ICT solutions monitor metrics that reflect improvements in the efficiency of these solutions [5]. Therefore, the analyses and comparisons of models involving different ICT solutions are prerequisites for the business planning processes since a timely application of relevant data represents an essential advantage within business modeling process. For a better overview of current trends in the adoption of ICT solutions, various data sets reflecting their adoption rates can be used [2]-[5]. Some conclusions that can be defined based on the analysis of these data are presented hereafter.

A) The generated network traffic and the number of Internet connections

An increase in the number of Internet connections is generally followed by a growth in total generated network traffic [4]. The main reason behind the existing massive network traffic growth is in the usage of audio and video-on-demand content, whose quality continuously increases [1]. Furthermore, the users' expectations related to the quality of services continue to rise. Uninterrupted high-speed connectivity is becoming an essential requirement for most users, regardless of their location or the chosen network access solutions. An increase in average traffic consumption per user is mostly caused by the rise in the adoption of bandwidth-intensive video services streaming. Besides the audio-visual media which accounts for the majority of network data traffic, the exchange of data traffic among end-user devices, terminal network equipment, servers and storage in the cloud continues to grow significantly, as well. Increase in types of available solutions users can choose to connect to the Internet, i.e., increase in the availability of Internet access technologies and services, follows growth in overall Internet network traffic demands. Intensive adoption processes of a wide range of Internet access solutions are currently available, as well [4]. The significant advances in the development of wireless ICT solutions have led to the massive adoption of mobile broadband connections [4]. Moreover, the growth of the mobile subscriber base is expected to continue within the next few years, with mobile broadband constituting a majority of the personal mobile subscriptions. Machine–to–Machine (M2M) connectivity also has the potential of becoming one of the contributors to the expected growth. The M2M services, as a part of the IoT solutions, which include automated communication and data transmission among two or more ICT entities, also record growth in demand [4]. Although common characteristics of IoT and M2M reside on remote access to devices, IoT is expanding the concept of M2M since it can be integrated into comprehensive and flexible business solutions. While IoT is focused more on software solutions and the Internet Protocol (IP) based networking, M2M communication is predominantly oriented on embedded hardware and mobile networks. However, considering the fact that M2M with Internet Protocol represents a part of IoT, the M2M/IoT services adoption trends can be analyzed jointly.

B) The generated network traffic and the package services

Considering the fact that every market has its limited maximal capacity, i.e., the total number of end-users, an additional increase in ICT market share can be achieved with growth in the number of adopted services. As already proven in many cases, services diversification represents a key driver of revenue growth for service providers, e.g., telecom operators or ICT enterprises, since end-to-end solutions enable the most significant differentiation. While some operators specialize in specific ICT segments, others focus on specific vertical markets. However, many operators tend to invest in new areas of growth. Therefore, the key to sustainable growth is in expanding beyond the core ICT services offerings. A possible solution for the increase in the number of services offerings can be achieved with the IoT services offerings. The IoT can induce positive growth because each offer can include a different combination of connectivity, devices, applications, and services. Many different types of ICT services offerings are currently available in the markets [5]. Although the stand-alone service offerings have a strong base of users, offerings of services packages comprising of more than one ICT service have also achieved fast adoption rates. These are any service packages where two or more ICT services are provided to users jointly, e.g., Internet, telephone services in the fixed network, telephone services in the mobile network or/and TV services. These are packages that jointly offer two or more services (e.g., 2D, 3D, and 4D packages). The 4D packages, which include Internet access, TV, and the fixed and mobile telephone services, note for fast adoption growth, mainly based on their total value [4]. This fact goes in line with the concept which suggests the creation of all-inclusive services offerings for the end-users, and a specific definition of the services' features [7].

C) The generated network traffic and the number of installed base stations

The seamless connectivity is one of the main contributors inducing growth in the adoption of wireless solutions, along with increasing bandwidth demands. The exponential growth of wireless network connectivity necessitates the convergence of dense heterogeneous networks since a single base station may not be able always to provide the high quality of services necessary for services demanding high data rates, as in a case of multimedia streaming.

Legacy mobile networks are dominated by macro cells served by high-powered cellular base stations whose radio coverage range of a few kilometers to tens of kilometers. As a response to the massive growth in network traffic, mobile operators have options to additionally upgrade their networks and provide even higher network capacities and user throughputs. One of the possible solutions for improvement is to maintain multi-standard radio access networks which provide capacity scalability due to increased spectral efficiency in existing bands. Also, improvements can be based on the usage of adequate modulation and multi-antenna techniques, and aggregation of a large number of licensed and unlicensed carrier bands. Furthermore, network densification also improves network capacity. It assumes changes in network topology or architecture by adding new cell sites. An appropriate network architecture should be chosen in combination with diverse factors. An increase in the number of radio sites using small cell sizes is an essential element for capacity increase. Small cells, as low-powered radio access nodes, are used to increase capacity without a need for tower-based radio sites. They operate in the licensed or unlicensed spectrum and typically cover areas range from ten meters to several hundred meters. Various types of small cells co-exist. These variants include femtocells, picocells, microcells, and metro cells. Network densification process achieved by adding small radio sites improves network coverage and capacity, enhances spectrum efficiency, and lowers energy power requirements. The usage of small cells operating on unlicensed bands is adequate in scenarios in which the deployment of network infrastructure is not commercially attractive for network operators, leaving that areas insufficiently covered by network services (e.g., as in some rural scenarios), as well as in scenarios with a scarce or limited volume of available network resources, as for instance, for a limited required radio frequency spectrum and a high network traffic demands (e.g. as in some urban scenarios). The small-cell based network access provides adequate coverage options for areas lacking basic network infrastructure (e.g., Fi-Wi or LPWA), as well as for offloading of network traffic on license-exempt frequencies (e.g., using Wi-Fi or LoRa) to free up the capacity in the macro network layer.

3. OVERVIEW OF COMPUTING NETWORK ARCHITECTURES

The geographic distribution represents the scale to which a system is widely spread or localized. Within this context, a type of network architecture that should be used in the given scenario, i.e., the centralized or decentralized architecture, depends mainly on the utility of implemented ICT system, and the intended usage of information extracted from the processing of collected data traffic. The available computing network architectures are the Cloud Computing (CC), Edge Computing (EC), Mobile Cloud Computing (MCC), Mobile Edge Computing (MEC), and Fog Computing (FC), as presented in Table I [8].

Table 1. Computing network architectures.

Network architecture	CC	EC	MCC	MEC	FC
Users	Any	Any	Mobile	Mobile	Any
Providers	Service providers	Enterprises / Network providers	Users / Service providers	Network providers	Users / Service providers
Initiative	Academic / Industrial	Academic / Industrial	Academic	Academic / Industrial	Academic / Industrial
Network architecture	Centralized / Hierarchical	Distributed / Localized	Central cloud & Distributed mob. devices	Localized / Hierarchical	Decentralized / Hierarchical
Internet connectivity	Necessary while running services	Not necessary, can operate autonomously	Necessary for offloading and obtaining content from the cloud	Not necessary, it can operate autonomously or connect to Int. using RAN	Not necessary, can operate autonomously
Hardware connectivity	WAN	WAN, (W)LAN, WiFi, cellular, ZigBee	WAN	WAN, cellular	WAN, (W)LAN, WiFi, cellular
Service access	Through the core network	At the edge	Through the core network	At the edge	Through devices from the edge to the core network

A) The Cloud Computing (CC)

A Cloud Computing (CC) model provides on-demand access to shared network computing resources [9]. Depending on the part of the application stack that can be managed by cloud users for processing, storage, and networking, the cloud offers the following models: infrastructure, platform, and software as services models (IaaS, PaaS, and SaaS) [10]. Because the demand for cloud resources can change within time, computing based on the provisioning of the required resources includes the virtualization for the deployment of on-demand applications. With the increase in the number of connected networking devices, services and applications, cloud architectures enable easy and cost-effective processes of computing, data caching and connectivity, but access to centralized resources can cause delays and degraded performance for devices that are located far from centralized cloud or data center sources.

B) The Edge Computing (EC)

Edge Computing (EC) architecture enables placing servers, applications, or small clouds at the edge of the network. The term ‘edge' used by the telecom operators usually refers to 4G and 5G base stations (BSs), Radio Access Networks (RANs), and Internet Service Providers' (ISP) access and edge networks. Moreover, that term is recently used also in the IoT context, as pointed in [11], and [12]. It refers to the local network in which sensors and IoT devices are placed. Therefore, the edge presents the first hop from the IoT devices, such as gateways or access points, but not the IoT nodes themselves. The usage of edge computing is intended to enable storage and compute resources closer to the user [13]. EC connects the IoT devices with the cloud. It enables data filtering, preprocessing, and aggregating using cloud services implemented near IoT devices [11].

- The Edge Computing (EC) vs. Cloud Computing (CC):

When placed at the network edge, storage, and compute systems reside closer to device, application, or user that produces the data to remove data processing latency. In this way, it is not necessary to send data from the edge of the network to some remote cloud or any centralized processing system, and back. By reducing the distance and time necessary to send data to the centralized system, the speed of data transfer, as well as the performance of services and applications on edge can be improved. The EC is adequate for industrial IoT usage cases since it brings processing closer to the sensors and actuators, and enables edge analytics of local data.

C) The Mobile Cloud Computing (MCC)

Mobile Cloud Computing (MCC), presents infrastructure outside of the mobile device where the data storage and processing are conducted [14]. MCC shifts most of the computation from mobile devices to the cloud, and therefore increases the mobile devices' battery life. However, offloading computation tasks to the cloud causes a relatively high latency for the delay-sensitive applications. MCC enables coordination between IoT devices, mobile devices, and cloud computing. This allows the running of data-intensive and computing-intensive IoT applications [15].

- The Mobile Cloud Computing (MCC) vs. Cloud Computing (CC):

MCC shares the characteristics of Mobile Computing (MC) and CC. As opposed to mobile computing which is resource-constrained, in MCC, the high availability of computing resources is present. In MCC, the availability of cloud services is higher than that of mobile computing.

- The Mobile Cloud Computing (MCC) vs. Edge Computing (EC):

Unlike MCC, EC is located at the edge of the network. Due to proximity to the IoT devices and users, latency in EC is in general lower than in MCC and CC. In the EC, connected devices are not limited by the resources as in standard mobile computing. EC uses small data centers and has higher service availability since devices do not have to wait for a centralized service.

D) The Mobile Edge Computing (MEC)

Within Multi-access Edge Computing, i.e., Mobile Edge Computing (MEC) system, functional, management, and orchestration entities, enable applications to run as virtual machines in a virtualized computing environment [16]. MEC elements are co-located with base stations. They deploy virtual machines for performing virtualization of routers and firewalls' functions to improve network efficiency, and cache content services to enhance user experience. Since edge architecture supports a specific access network, either wireless or wireline, the MEC infrastructure is deployed and owned by the telecom operators.

- The Mobile Edge Computing (MEC) vs. Cloud Computing (CC):

MEC is an extension of MC through EC. MEC presents a platform providing CC features within the Radio Access Network (RAN) close to mobile users.

- The Mobile Edge Computing (MEC) vs. Edge Computing (EC):

MEC extends EC by enabling computing and storage near mobile devices. MEC enables adding of EC functionality to the existing RAN base stations. In MEC, small data centers with virtualization can be used. In MEC and EC, computing resources are lower than in CC due to the available hardware. MEC supports low-latency applications. Both EC and MEC can operate even without Internet access. While MEC enables connections through a WAN, cellular, or WiFi, EC enables connections using LAN, cellular, or WiFi. MEC enables EC to various mobile devices [17], as well as the usage of applications sensitive to delays over the mobile network [18]. MEC has also incorporated the Software-Defined Networking (SDN), as well as Network Function Virtualization (NFV) capabilities. SDN enables easy management of virtual networking devices through software Application Programming Interfaces (APIs), and NFV enables faster deployment of networking services through virtualized infrastructure [19]. Using SDN and NFV, the orchestrator can be used to coordinate the resource provisioning across multiple network layers [20].

- The Mobile Edge Computing (MEC) vs. Mobile Cloud Computing (MCC):

Increased adoption of wireless solutions induces further growth in mobile data traffic. The generated large quantities of multimedia traffic need to be processed by keeping up to demands set on users' experience. To overcome the existing data processing limitations of radio access networks, the following complementary approaches are proposed: one which suggests the centralization of base station functions using virtualization and shifting of computing capabilities to the central cloud, i.e., the Cloud Radio Access Network (C-RAN), and the other which suggests shifting of computing capabilities to the edge, i.e., the Mobile Edge Computing (MEC). Unlike MCC, related to the cloud service users of mobile devices and cloud service providers, MEC focuses on RAN-based infrastructure [18].

E) The Fog Computing (FC)

In Fog Computation (FC), storage, computing, and data management occur in the cloud but also along the path on which data travel to the cloud. FC presents a horizontal architecture platform that enables computing, storage, control, and networking functions closer to the users [21], and allows the distribution of computing functions between different platforms [22]. In FC devices either serve as computing nodes or use fog resources. FC is mainly implemented in devices (e.g., small servers, gateways, access points, routers, or switches) owned by ICT enterprises.

- The Fog Computing (FC) vs. Cloud Computing (CC):

While CC must be accessed using the core network, FC can be accessed using connected devices from the edge to the network core. While CC provides computing resources using high power consumption, FC provides computing resources at lower power consumption [23]. CC devices need Internet connections for cloud services. The FC can work independently and send necessary updates to the cloud when an Internet connection is available.

- The Fog Computing (FC) vs. Edge Computing (EC):

FC extends EC capabilities given computation distribution and traffic load balancing. While EC orchestration and management derive from specific vertical practices of legacy systems, such as mobile network, FC provides an architecture which incorporates tools for distributing, orchestrating, and managing resources and services across networks. FC orchestration enables the pooling of resources and collaborations between fog nodes which helps load balancing. FC and EC both move the computation and storage to the network edge. However, while FC provides computing, networking, and storage in any place from cloud to devices, EC provides computing at the edge [21]. EC is optimized for a single type of network resources. FC supports cooperating nodes running distributed applications, and heterogeneous environments on any node. FC's architecture permits every fog node to be equipped with the necessary dynamically configured resources and provides a balance of computation and storage capabilities. While FC focuses on interactions between edge devices, EC focuses on the technology of connected devices [12].

- The Fog Computing (FC) vs. Mobile Cloud Computing (MCC):

FC can be integrated within the radio access networks (RAN), and form the so-called Fog RAN (F-RAN). F-RAN may be used for data caching at the edge [24]. Cloud RAN (C-RAN) virtualizes the base station functions [25] and provides centralized control over F-RAN nodes. Both F-RAN and C-RAN are appropriate for cellular networks with base stations.

- The Fog Computing (FC) vs. Mobile Edge Computing (MEC):

The MEC computing process aligns with the emerging concept of FC. However, these somewhat differ. While MEC extends computing capabilities to the edge of the radio access network with a new interface between the base stations and upper layers, FC architecture brings the processing and storage resources to the lower layers for occasionally connected mobile ad-hoc and sensory devices.

The majority of the research conducted in the field of edge and fog computing is related to schemes that deal with the topics related with improvement of the Quality of Service (QoS) by minimizing latency or data losses, the topics related with the scalability by efficiently scaling to the large magnitude of IoT networks, and the topics related with the heterogeneity of devices, as presented for instance in [26], and [27]. Also, the network management schemes are in the research focus, as presented, for instance, in [28] and [29]. The problems related to management of latency-sensitive IoT applications is evident in [30]. Moreover, it is important to mention that the cloud applications can fully or partially migrate to edge [31]-[32], and inversely [33]-[34] however the streaming a massive amount of data to the cloud imposes considerable energy consumption.

4. PREDICTIVE MODELS

An overall increase in generated network traffic and a limited amount of available network resources have encouraged telecom operators and enterprises working within ICT domain to search for the optimal network architectures to enable the best approaches for storing and management of generated data traffic, applying analytics over gathered data, and deriving useful knowledge. Within this context, forecasting of ICT solutions' adoption rates can be used within planning processes to achieve efficient distribution and allocation of available network resources. The forecasting of adoption rates of various ICT solutions is increasingly important in optimal resources management.

Different predictive models whose implementation can impact planning processes are used [35]. The processes used for the selection of the forecasting methods are described in [36]. In processing time-series data, one of the most commonly used methods includes data classification. There are many examples of usage of data classification processes, some of which are applied, for instance, in adapting the mobility management mechanisms [37], prediction of applications' data consumption [38], and user activity [39].

In this paper, several commonly used models for time series data analytics, described for instance in [40] and [41], and some additional models, described in more detail in [42], are taken into consideration. In [42], the analyses are conducted to point to the fact that the presented models enable adequate forecast of the number of future service users, and in [41], to point to the fact that these models allow adequate forecast for finding the best service offerings. However, the aim of the analyses conducted in this paper, unlike the ones conducted in [41] and [42], is to point to the fact that predictive modeling processes can be used for selecting optimal network architectures for storing and processing of exponentially increasing generated network traffic, as well. This is particularly important for enhancing data processing speed and achieving higher levels of quality of services.

A) Common Models Used in Predictive Modeling

The scope of this paper covers the prediction-based analyses of ICT solutions adoption, used as the starting point to further processes of network resources distribution and allocation planning. The several standard, i.e., commonly used [40], as well as some additional predictive models [38] are used, to show and compare their predictive accuracy. The common models for the forecasting of adoption, and their related expressions for the Simple Logistic model, Richards model, Bass model, and Gompertz model, respectivelly, are:

$$L(t;M,a,b)=\frac{M}{1+e^{-a(t-b)}} \tag{1}$$

$$R(t;M,a,b,c)=\frac{M}{\left[1+e^{-a(t-b)}\right]^{c}} \tag{2}$$

$$B(t;M,p,q,t_s)=M\cdot\frac{1-e^{-(p+q)\cdot(t-t_s)}}{1+\frac{q}{p}\cdot e^{-(p+q)\cdot(t-t_s)}} \tag{3}$$

$$G(t;M,a,b)=M\cdot e^{-e^{-a(t-b)}} \tag{4}$$

where L, R, B, and G represent the volume of adopted solutions over period t, determined using the Logistic, Richards, Bass, and Gompertz models, respectively. The following parameters define these models: M, which reflects the market capacity; a, which indicates the speed of adoption; b, which positions the graph on the timescale; and c, which places the model's inflection point; p, which reflects the coefficient of innovation ($p> 0$); q, which demonstrates the coefficient of imitation ($q \geq 0$); and t, which reflects the time when the solution was introduced in the market ($t\geq ts$).

B) Additional Predictive Models

To expand the analysis and compare features of additional models, combinations of some other parameters are taken into consideration, and combined models are derived, as described in more detail in [38], using the following expression:

$$\boldsymbol{B}\ \ (t) = M \cdot \frac{e^{\left[1-e^{-a(t-b)}\right]^{d}}}{e^{\left[1+e^{-a(t-b)}\right]^{c}}} \tag{5}$$

where $BB(t)$ denotes the volume of adopted solutions, and M a total capacity.

Table 2. Overview of additional predictive models.

Models:	Parameters values: Parameter c:	Parameter d:	Notes:
Logistic (L)	1	0	
Bass (B)	1	1	
Richards (R)	$c\epsilon\left|0,+\infty\right\rangle$	0	For c=1: R ≡ L
Gompertz (G)	0 1	1 0	Subcases of c for d=0 and d=1
GB	1	1	
GR	$c\epsilon\left|0,+\infty\right\rangle$	0	Subcases: (c=0, d=0) and (c=1, d=0)
GBR	$c\epsilon\left|0,+\infty\right\rangle$	1	Subcases: (c=0, d=1) and (c=1, d=1)

These modified forms take into consideration several additional combinations of parameters' values, previously defined in [38], as presented in Table III.

$$GB(t; M, p, q, t_s) = M \cdot \frac{e^{\left[1-e^{-(p+q)\cdot(t-t_s)}\right]}}{e^{\left[1+\frac{q}{p}\cdot e^{-(p+q)\cdot(t-t_s)}\right]}} \tag{6}$$

$$GR\ (t; M\ , a, b, c) = M\ \cdot \frac{e}{e^{\left[1+e^{-a(t-b)}\right]^{c}}} \tag{7}$$

$$GBR\ \ (t; M\ , a, b, c) = M\ \cdot \frac{e^{\left[1-e^{-a(t-b)}\right]}}{e^{\left[1+e^{-a(t-b)}\right]^{c}}} \tag{8}$$

These models combine the features of the Gompertz (G), Bass (B) and Richards (R) models. They model the fast growth and are determined by the same parameters, M, a, b, c, p, and q, as common models. The predictive models can be additionally modified using more explanatory parameters. Although certain generalizations of the existing models expand their features' description, additional parameters require larger sets of known data points used in the predictive modeling process, which limits their usage.

5. MODELING OF ICT SOLUTIONS ADOPTION

The given models (1)-(8) are used in forecasting adoption trends of several ICT solutions. The analyses of the accuracy of fitting and forecasting processes are conducted, and the parameters estimated within the fitting processes are used to generate the forecasts of future values, based on the known ones. The chosen data sets comprise data reflecting the total mobile and fixed network data traffic [2]-[4], total wireless data transmission capacity across all frequency bands [3]-[4], total number of GSM, UMTS and LTE base stations [3]-[4], number of users of stand-alone Internet services [5], number of users of 4D service packages [4]-[5], and number of users of the M2M/IoT services [4]-[5].

A) Fitting Process

As can be seen from the Figures 1-6, the fitting processes comprise the adjustments of models' parameters to best describe the real time series values (denoted as 'Data' [2]-[5]), representing the trends in adoption of several chosen ICT indicators. The fitting process is conducted within the time period from 2011 to 2020 in order to point to the fact that the smaller set of known data values is sufficient for forecasting of further values. The presented results point to the fact that the accuracy of the Bass model improves if the number of known data points, i.e., the ones used for training, starts with lower values. All other common models and GR model show good fitting properties in the given cases.

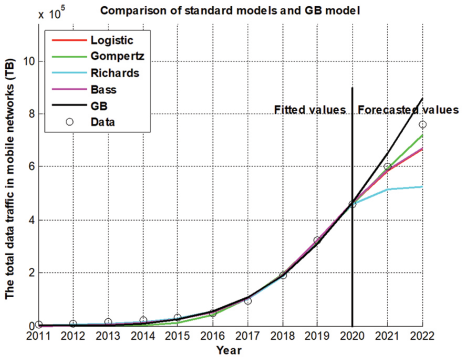


Fig. 1. The total data traffic in mobile networks (TB) [2]-[5]

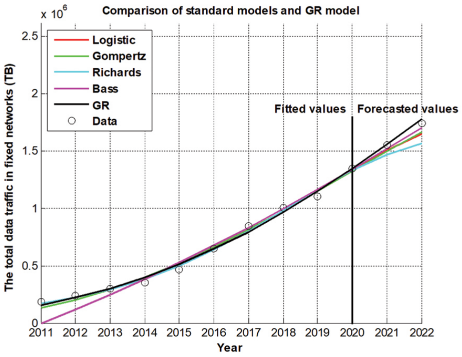


Fig. 2. The total data traffic in fixed networks (TB) [2]-[5]

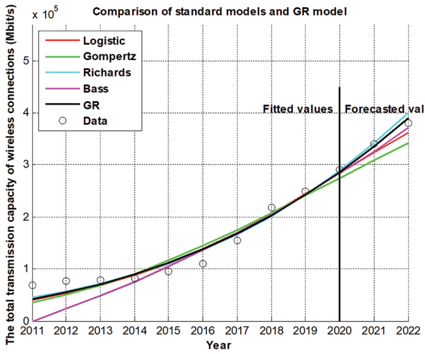


Fig. 3. The total transmission capacity of connections across all frequency bands (Mbit/s) [3]-[4]

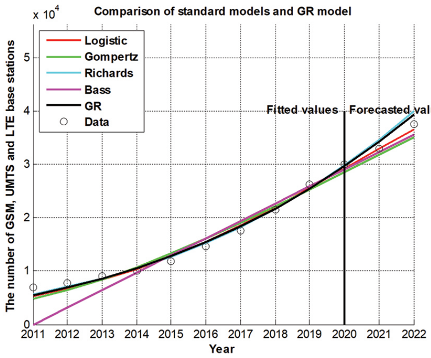


Fig. 4. The number of GSM, UMTS, LTE base stations [4]

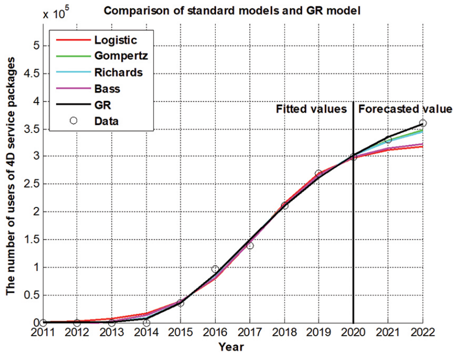


Fig. 5. The number of users of 4D service packages [3]-[5]

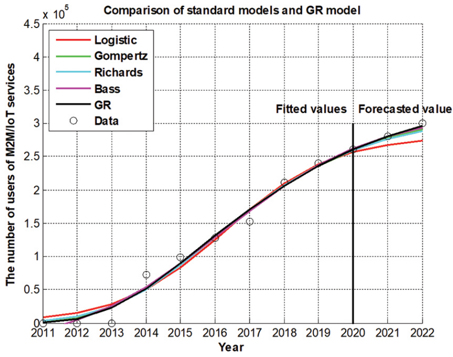


Fig. 6. The number of users of the M2M/IoT services [3]-[5]

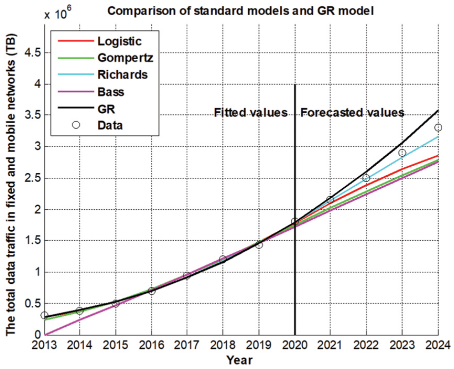


Fig. 7. The total data traffic in fixed and mobile networks (TB) [2]-[5]

Moreover, these models show very good fitting properties used for the longer forecasting time periods, as well, as presented by the values shown in Figure 7.

In Figure 7, the real data for the period 2014-2020 [4]-[5], and assumed data values of further growth in the period 2021-2024 are used according to the existing total network traffic growth trend.

B) Forecasting Process

Several measures are used to determine the accuracy of conducted forecasts. Statistical criteria can be selected after deciding on the general type of forecasting method [36]. There are mainly four types of forecast-error metrics: scale-dependent, percentage-error, relative-error, and scale-free error metrics. The chosen statistical parameters that describe the accuracy of forecasted time series values are the forecast error and the mean absolute deviation. These are adequate metrics in analyzing the error for a single output and considering the fact that the prediction errors are in the same unit as the original series. The Mean Absolute Deviation (MAD), also commonly called the Mean Absolute Error, is the measure of aggregate error defined by the expression:

$$MAD = \frac{\sum_{i=1}^{n} |E_i|}{n}$$

where n is the number of prediction errors which are used for the calculation, and E, forecast error, i.e., the difference between the actual value and the forecasted value in the corresponding period t. A smaller amount of the mean deviation denotes the model's better prediction performance.

The sample data set used for modeling is divided into subsets which comprise the training data (the shaded ones in tables below Figures 8-14) - used for the model parameters fitting, and the testing data (all other) - used for determination of the accuracy of the forecasted values. To determine the accuracy of the forecasted values, not only training data, but also testing data must be known, so data from the reports [2]-[5] are used.

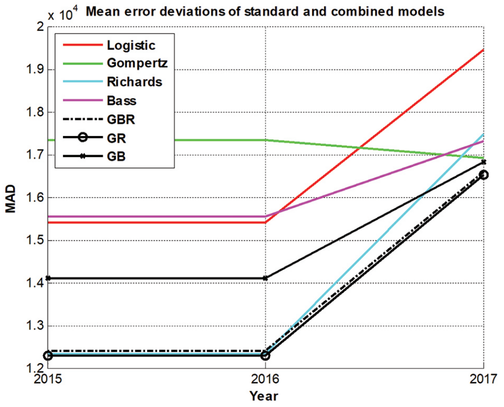
Considering the gathered results of the conducted fitting processes presented in Figures 1-7, and the undertaken forecasting processes shown in Figures 8-14, it can be concluded that the primary difference among the models' fitting and forecasting accuracy is caused by different positions of the models' inflection points.

As presented in Figures 1-14, the Bass model shows limitations both in fitting and in the forecasting of the initial short-term upper market capacity. All other models show very good fitting properties, as presented in Figures 1-7. Moreover, the common models show adequate accuracy in forecasting, as well, as presented in Figures 8-14.

The additional GB, GR and GBR models combine the features of the Gompertz (G), Bass (B), and Richards (R) models. The combined models that have the features of the Gompertz model accurately predict the fast growth. However, the lack of the Gompertz model relates to the fact that it cannot limit the excessive increase in the long run, and this can reflect the forecasting accuracy of the combined models, as well. Moreover, since the Bass model has difficulties in assessing the exact upper market capacity limit in the initial growth phase, the forecasting accuracy of the Bass model combined solely with the Gompertz model, i.e., the GB model, is also not always adequate, as presented in Figures 8-14. However, the model that combines the features of the Bass model with the Gompertz and Richards models, i.e., GBR model, is more accurate for forecasting of the long- term adoption of the services since having a flexible inflection point which limits the accelerated growth in values, as presented in Figures 8-14. The combined models that use the features of the Richards model, i.e., the GR and GBR models, generally show very good forecasting properties even if the minimum number of values is used in fitting, as presented in Figures 8-14. The Richards model accurately forecasts significant growth in the long run since it uses a flexible inflection point to adjust growth to the last existing training value, which can be seen for the GR and GBR models, as presented in Figures 1-14.

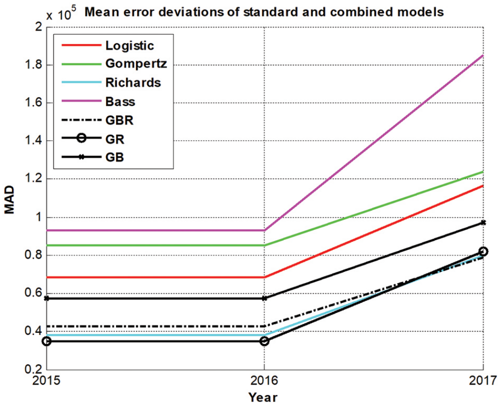
For a sum-up of the presented results, the models that combine the features of the Richards model with the Gompertz model achieve proper fitting to fast growth and show very good forecast results in all presented cases.

The purpose of the conducted analyses is to point to the fact that the predictive models can be used to adequately forecast values in many different usage cases, for instance, in the case of expected further growth in the total (fixed and mobile) network data traffic, as presented in Figure 7, as well as ICT indicators associated with it - growth in the network bandwidth usage, increase in the number of base stations, changes in the number of users of stand-alone services, as well as the growth in the number of users of package services and the M2M/IoT services. Moreover, additional significance and usefulness of these given forecasts are presented hereafter.



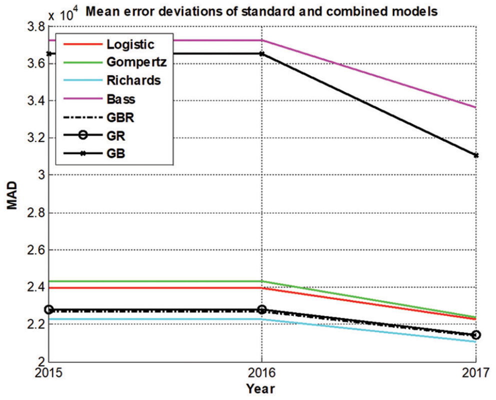
Period:	2011	2012	2013	2014	2015	2016	2017
The total data traffic in mobile networks:	3.552	8.329	15.712	22.270	32.408	49.173	92.033

Fig. 8. The total data traffic in mobile networks (TB) [2]-[4]



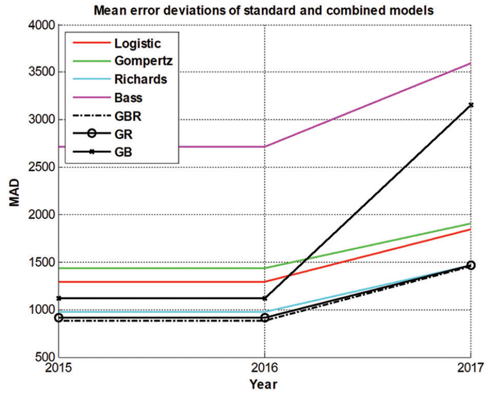
Period:	2011	2012	2013	2014	2015	2016	2017
The total data traffic in fixed networks:	188.821	238.408	299.890	362.482	480.880	660.547	846.846

Fig. 9. The total data traffic in fixed networks (TB) [2]-[4]



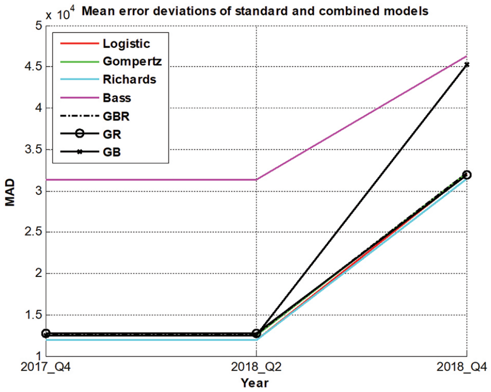
Period:	2011	2012	2013	2014	2015	2016	2017
Internet bandwidth usage (Gbit/s):	69.000	77.000	79.000	82.000	95.000	110.000	155.000

Fig. 10. The total transmission capacity of connections across all frequency bands (Mbit/s) [3]-[4]



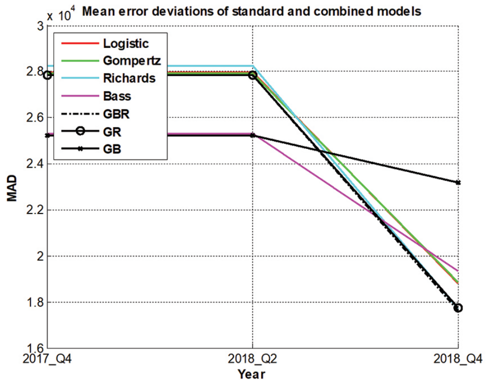
Period:	2011	2012	2013	2014	2015	2016	2017
The total number of GSM, UMTS, and LTE base stations:	7.031	7.709	9.026	9.985	11.914	14.711	17.566

Fig. 11. The number of GSM, UMTS, and LTE base stations [3]



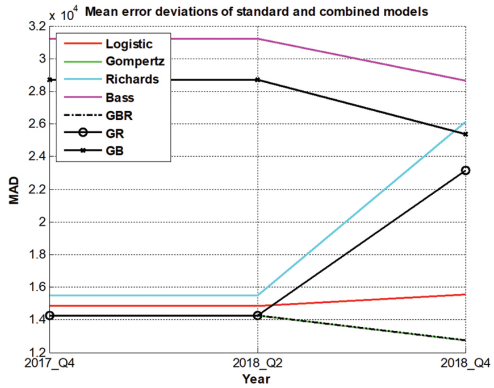
Period:	2015_Q4	2016_Q2	2016_Q4	2017_Q2	2017_Q4	2018_Q2	2018_Q4
The number of users of 4D packages:	35.772	79.168	96.750	98.642	138.536	166.807	211.762

Fig. 12. The number of users of 4D service packages [3]-[5]



Period:	2015_Q4	2016_Q2	2016_Q4	2017_Q2	2017_Q4	2018_Q2	2018_Q4
The number of users of stand-alone Internet services:	178.046	130.932	139.599	191.709	165.415	186.628	136.454

Fig. 13. The number of users of stand-alone Internet services [5]



Period:	2015_Q4	2016_Q2	2016_Q4	2017_Q2	2017_Q4	2018_Q2	2018_Q4
The number of users of M2M/IoT services:	98.375	108.138	127.446	147.716	152.654	172.142	211.677

Fig. 14. The number of users of the M2M/IoT services [4]-[5]

C) Analysis of results

Advanced telecom networks implement features that allow simultaneous management of network traffic originating from various types of terminal devices, services, and applications, having different requirements on the speed of data processing. The following analysis is conducted to point to the fact that the methods used in the selection of adequate network architectures for optimal distribution and allocation of available network resources and efficient data processing should be based on the forecasts of additional network traffic growth given by the best forecasting methods. According to the forecasted trends in growth of the network traffic presented in Figure 7, the forecasts of growth in adoption of related ICT solutions can be assumed as well, which are also assumed in reports [1], [4] and [6]. The processes of distribution and allocation of network resources used for data caching and computing based on predictive analysis results are placed in the scope of conducted research. The following analysis considers applying forecasted trends in adoption rates of ICT solutions as a starting point in the planning of distribution and allocation of network resources, as well as in the selection of adequate network architectures. The following guidelines can be defined based on the previously presented results, and reports [1]-[6].

- The results of the conducted analyses presented in Figure 1, Figure 2 and Figure 7 and reports [1], [4] and [6] forecast further growth in mobile and fixed network traffic.

The increase in data traffic is directly related to the development of the digital society. A significant increase in network data traffic is visible in both fixed and mobile networks, especially in the fixed network [4]. Moreover, within the last four years, the increase in data traffic has also been recorded among broadband users accessing Internet via wireless technologies in a fixed network [4]. Considering the importance of broadband internet access, as well as expected further investments in fiber access networks and 5G technology [4], support for significant additional growth of data traffic is expected in the following years.

Development in fixed communication networks is going in the direction of high availability of ultra-fast fiber optic networks, and in mobile networks in the direction of the introduction of the new 5G technologies [4]. The increasing convergence of these networks in the future will not only lead to even greater availability and the quality of existing services than to the emergence of new services and business models [6].

European policy makers have set broadband connectivity targets for Europe, and both wired, notably fiber, and wireless technologies play important role in delivering the target. The directives of the European Parliament and Council 2014/61/EU, refer to fixed (wired) and wireless to lower the costs for deploying broadband, while several national broadband development plans explicitly acknowledge the role of FWA, as a combination of different fixed and wireless technologies [6].

Table 3. Features of computing network architectures

Forecasts:	Demands:	Necessary features:	CC	EC	MCC	MEC	FC
Increase in generated network traffic	Content offloading using complementary network solutions	Heterogeneity support	✓	✓	✓	×	✓
	Content offloading using additional servers	Distributed storage support	×	✓	×	✓	✓
	Computation offloading	Virtualization support	✓	✓	×	✓	✓
Increase in bandwidth usage for wireless traffic	Network densification using small cells	Access to services not necessary through the core network to approach centralized server	×	✓	×	✓	✓
Increase in usage of multimedia package services	Multimedia streaming	Ultra-low latency support	×	✓	×	✓	✓
		Real-time applications support	×	✓	×		
Increase in usage of IoT services	IoT ecosystem deployments	Multiple IoT applications and devices support	✓	✓	✓	✓	✓

In report [6], in densely populated areas of larger cities funding of fixed solutions used within FWA is considered unnecessary with the deployment of 5G access. However, in larger city industrial zones, as well as on highways along which the wired (fiber) infrastructure is implemented, additional funding of fixed infrastructure upgrade can be adequate solution even for the FWA models that can be used to provide both ultra-high capacity and mobility for the real-time communication, especially in the cases of smart environments and autonomous driving, having very high network traffic requirements. In these scenarios, upgrade of fiber infrastructure should be additionally funded by EU funds in order for FWA infrastructure to give its best possible application results, and to enable sufficient capacity and support for high traffic requirements related to autonomous driving. In these cases, it is possible to lease fixed network access to mobile operators to achieve convergence of ultra-high fixed capacities and necessary mobility, also using the converged FWA implementation business model.

Considering the given forecasts which point to further growth in generated mobile and fixed network traffic, the guidelines related to effective busy hour and real-time network traffic management can be additionally defined concerning the deployment of content offloading processes (used in CC, EC, MCC, MEC and FC computing network architectures) or computation offloading processes (used in CC, EC, MEC, and FC computing network architectures), applied to increase data processing speed and improve users' quality of experience with a shorter delay. In general, for the majority of network traffic, there is no added value to route the data through the core network. In this case, the offloading process can be carried out. The content off-loading can be achieved by switching the traffic to use complementary network technologies for delivering data, freeing bandwidth and reducing the total amount of data being carried over a particular communication channel, but also allowing the selection of adequate communication channel for better connectivity. Either the client or operator can set the rules triggering the off-loading action. It is possible to select traffic off-loading at different locations, over open or secured license-exempt access links, and depending on the demanded quality of service. Furthermore, the mobile edge (EC) and fog (FC) based networking architectures enable distribution of data storage at the edge of the network and in that way enable better data processing efficiency and reduced latency to users. Although edge systems scale by adding more resources at a given location, for instance, the small clouds, this approach is not adequate for scaling to support the massive number of devices. Fog system is, on the other hand, capable of shifting computation, networking or storage tasks across peer nodes, or between the cloud and fog, and enables resources pooling. Moreover, the recent developments in mobile edge and fog computing concepts are leveraging small cells as possible computing platforms. Edge computing uses virtualization to distribute computing resources locally. Fog architecture additionally extends edge capabilities by supporting hardware virtualization at each node and allows data processing to be moved to adjacent nodes if some node in the network is unavailable or overloaded. However, although offloading processes increase the efficiency of data processing, they do not always manage to reduce overall systems' capacity consumption. Due to the significant expected increase in the traffic demands, the volume of network resources necessary to achieve a defined level of quality of service also increased. In this context, the selection of effective offloading strategies, and also network infrastructure upgrade must be carried out.

- The results of the conducted analyses presented in Figure 3 forecast further growth in transmission capacity of wireless connections across all frequency bands (bandwidth) usage, as well as growth in the number of installed base stations presented in Figure 4.

According to market indicators given in [4], a continuous growth in demand for the broadband internet access services is present. To meet the increased service demand while maintaining the level of service quality, it is necessary to increase network capacities and access speeds, i.e., to invest in high-speed and high-capacity access networks.

In larger cities, the base station inter-site distance is supported by the need to provide capacity rather than range. This assumes co-location by all operators but in practice there are likely to be many more small cells sites because not all sites will have colocation [6].

Considering the given forecasts which point to further growth in wireless bandwidth usage presented in Figure 3 due to expected significant growth in generated wireless network traffic [4], the guidelines related to effective bandwidth management can be defined concerning network densification processes. The overall increase in bandwidth usage presents the main driver to deploy small cells and to justify the further network densification [6], since the denser networks imply deployment of more edge-oriented services closer to the users (EC and MEC), and this greatly improves the quality of user experience (as well as FC solutions' application) and reduces the network traffic loads on the backhaul links. To identify the point at which small cells become necessary to supplement macro cellular networks, the traffic volume density per allocated unit of bandwidth (Gbps/km2/Hz) is used as the metric. Network densification starts by deploying small cells when the parameter exceeds the 0.02 Gbps/km2/Hz threshold. Furthermore, the need for small cells will be even more necessary in the next generation network settings since the higher spectrum bands need denser network deployments to support larger traffic volumes [6]. While in 4G/LTE networks site densities of up to 30 sites/km2 are common, a 5G network densification process assumes the ultra-dense networks with site deployment densities of 90 sites/km2. According to data presented in [4], the reported number of UMTS and LTE base station sites in 2017 reached 12.440 sites in total. This reflects the joint average site densities of 0,21 sites/km2 in UMTS/LTE networks, which implies that there is a possibility for further network densification, also presumed within the 5G networking concept. This is also visible from the significant growth in the volume of the generated network traffic, i.e., from the fact that the mobile network data traffic for example in 2018, compared to 2017, showed an increase of 103%, as reported in [5], and that further significant network traffic growth is assumed within the next several years [4], as presented in Figure 7. The network traffic growth rates will be even higher once smart environments and solutions become implemented extensively, as pointed out also in [1] and [6].

- The results of the conducted analyses presented in Figure 5 forecast further growth in the number of users of package services.

Given the data from [3] and [5], the operators offer their services to end users in service packages much more than independently, as presented in Figure 9 and Figure 10. The significant increase in the number of 4D packages, i.e., service packages in which operators offer customers in one account services in the fixed and mobile network, is visible. This is possible for fixed network operators that are convergent operators, i.e. the operators that with fixed services can also offer services in mobile networks. The number of users of 4D packages is growing, while the number of users of 2D and 3D packages [4], as well as the number of users of stand-alone Internet services [5], is reducing.

Considering the given forecasts which point to further growth in the number of users of package services demanding the transfer of heterogeneous data traffic and multimedia streaming, the guidelines related to the effective management of large amounts of multimedia content can be defined concerning the deployment of distributed storage systems (EC and MEC). Unlike in the case of the number of users of stand-alone services, the forecasts point to significant growth in the usage of multimedia package services. The large quantities of multimedia event streams need to be efficiently processed. To keep up with demands set on users' experience, and to overcome the limitations of current radio access networks, the emerging context suggests moving computing capabilities to the edge, as well as the usage of fog computing (FC). The MEC servers implemented directly at the base stations allow faster computing, high-volume media content storage, and hosting compute-intensive applications close to terminal devices. In this way, the MEC systems bring various network improvements, such as the fulfillment of the ultra-low latency requirements, pre-processing of large volumes of data before data forwarding to the cloud, and context-aware services information.

- Finally, the results of the conducted analyses presented in Figure 6 forecast further growth in the number of users of M2M/IoT services.

The traffic demand from non-human usage is just at the beginning of its growth curve [4]. Therefore, new use cases need to be considered. According to [6], connected cars, cameras, and a high density of IoT devices in smart environments will generate significant amounts of new data traffic. Traffic generated by connected vehicles, cameras, and video-based sensors could be a multiple of traffic generated by human users [6].

Considering the given forecasts which point to further growth in the number of users of M2M/IoT services, the guidelines related to the management of network resources within the IoE ecosystems can be defined concerning the deployment of fog and edge solutions. Concerning the ambient intelligence of computing and sensory devices embedded in the environment, smart environments are generating significant quantities of data, and this induces progress towards next-generation data-intensive intelligent systems. With remote sensors installed on machines, components, or devices, different types of data are generated. Concerning the management of limited available network resources and a need for large-scale data classification and clustering processes, the heterogeneous data-rich ecosystems present new challenges in designing intelligent systems. IoT devices have limited computational, memory, and energy resources, so they heavily rely on edge (EC, MEC) and core networks for data handling, processing, and analysis. If data is sent back across a long network link to be analyzed, logged and tracked (CC), that process takes much more time than in the case in which the data is processed at the network edge, close to the source of the data (EC, MEC and MCC). The fog (FC) and edge systems (EC and MEC) enable better options for IoT users and technology providers. By removing the limits imposed by centralized network architectures based on data processing on centralized cloud servers (CC) allows deployments of more distributed and flexible IoT systems.

6. CONCLUSION

Since novel ICT technologies, applications, and services bring many advantages to end-users, ranging from smart homes and smart cars, to smart factories and smart environments, further growth in the adoption of these solutions, as IoT solutions, is inevitable. However, given the accelerated growth in the volume of generated network traffic which is closely correlated to the adoption of advanced ICT solutions, some of the main challenges telecom operators and enterprises working within ICT domain currently cope with are related to the enabling of efficient processing of large amounts of data. Valuable information extracted from a large volume of collected data presents added value that encourages operators and enterprises to experiment with the implementation of novel ICT solutions and to invest in deployments of large-scale IoT initiatives. Therefore, available networking architectures and network settings that enable efficient distribution and allocation of available network resources used for data caching, processing, and computing are examined. Due to requirements related to the performance of novel ICT systems, optimal selection of adequate networking architectures can be achieved based on the analysis of data usage patterns. The processes of effective distribution and allocation of network resources, which significantly impact data processing speed, processing latency and energy consumption can be supported by adaptation and upgrade of existing network architectures.

Therefore, it is necessary to carefully consider all aspects of justification of avoidance of usage of certain funding schemes for network upgrade, proposed for instance in the report analyzing 5G FWA implementation scenarios within [6], and it is important to suggest valid FWA scenarios in which funding schemes for fixed network upgrade are justified and desirable, with regard to their specifics.

In this paper, the analytics of the adoption processes of different telecommunications solutions is conducted for the gathered sets of time series data. The predictive modeling process of broadband services adoption using several types of adoption growth models are given. Alongside standard predictive models, some additional models are used to extend the analysis and additionally back up the results collected by commonly used models. The prediction-based processes of planning, distribution, and allocation of network resources present a crucial step to achieve effective resource planning and network management and to improve the overall system performance. The presented results point to further fast expected growth of overall generated network traffic. Although the telecom operators and ICT enterprises' operation contexts differ and their demands on network infrastructure features may be somewhat different, the guideline that can be drawn from the collection of presented traffic growth forecasts and market estimates includes the fact that the existing network architectures should be developed towards edge and fog networking concepts to enable efficient processing of large amounts of generated data.

It is important to emphasize that the analyses are conducted only for a particular available data sets since a very reduced amount of statistical data is publicly available. In the cases when additional data sets are available, the more precise planning processes can be achieved using the same presented forecast-based approach. Considering the fact that the choices related to network design can be based on more than one parameter, multi-objective decision approaches are planned in future research work.

7. REFERENCES:

[1] "IMT traffic estimates for the years 2020 to 2030", Report, Mobile, radiodetermination, amateur and related satellite services, ITU-R M.2370-0, 07/2015

[2] Croatian Regulatory Authority for Network Industries, "Yearly Work Report for 2017", 2018, https://www.hakom.hr/UserDocsImages/2018/izvjesca_i_planovi/Godišnje_izvješće_HAKOM_za_2017.pdf?vel=3172981 (accessed: 2021)

[3] Croatian Regulatory Authority for Network Industries, "Yearly Work Report for 2018", 2019, https://www.hakom.hr/UserDocsImages/2019/izvjesca_i_planovi/HAKOM_GI_2018.pdf?vel=16546641 (accessed: 2021)

[4] Croatian Regulatory Authority for Network Industries, "Yearly Work Report for 2019", 2020, https://www.hakom.hr/UserDocsImages/2020/izvjesca_i_planovi/HAKOM%20GI2019%20HR%2020200623.pdf?vel=3616097 (accessed: 2021)

[5] Quarterly Data Reports 2011-2020, Croatian Regulatory Authority for Network Industries, "Quarterly Comparative Data for the Electronic Communications Market in the Republic of CroatiaData", 2021, https://www.hakom.hr/?id=6426 (accessed: 2021)

[6] Report by Coleage Consults Ltd., GSMA, 12/2020, "IMT Spectrum Demand: Estimating the mid-term bends spectrum in the 2025-2030 timeframe, https://www.gsma.com/gsmaeurope/resources/imt-spectrum-demand/

[7] S. Moyer, "Networked Appliances: The Next Wave of Computing?", Proceedings of the 17th International Workshop on Feature Interaction in Telecommunications and Software Systems, Ottawa, Canada, 2003.

[8] A. Yousefpour et. al., "All One Needs to Know about Fog Computing and Related Edge Computing Paradigms", A Complete Survay, Vol. 98, 2018, pp. 289-330.

[9] P. Mell et al., "The NIST Definition of Cloud Computing", 2011

[10] T. Dillon, C. Wu, E. Chang, "Cloud Computing: Issues and Challenges", Proceedings of the IEEE International Conference on Advanced Information Networking and Applications, Perth, WA, Australia, 20-23 April 2010.

[11] A. Reale, "A Guide to Edge IoT Analytics", International Business Machines, https://www.ibm.com/blogs/internet-of-things/edge-iot-analytics (accessed: 2021)

[12] "What is Edge Computing?", General Electric, https://www.ge.com/digital/blog/what-edge-computing (accessed: 2021)

[13] Technical Report, Open Edge Consortium, "About - the Who, What, and How", http://openedgecomputing.org/about.html (accessed: 2021)

[14] H. T. Dinh, C. Lee, D. Niyato, P. Wang, "A Survey of Mobile Cloud Computing: Architecture, Applications, and Approaches", Wireless communications and mobile computing, Vol. 13, No. 18, 2013, pp. 1587-1611.

[15] Technical Report, National Institute of Standards and Technology, "Mobile Cloud Computing", https://www.nist.gov/programs-projects/mobile-cloud-computing (accessed: 2021

[16] F. Giust et al., "MEC Deployments in 4G and Evolution Towards 5G", ETSI White Paper, 2018

[17] T. Taleb et al., "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration", IEEE Communications Surveys and Tutorials, Vol. 19, No. 3, 2017, pp. 1657-1681.

[18] Y. Chao Hu et al., "Mobile Edge Computing – A Key Technology Towards 5G", ETSI white paper, Vol. 11, 2015, pp. 1-16.

[19] K. P. Kadiyala, J. A. Cobb, "Inter-As Traffic Engineering with SDN", Proceedings of the Conference on Network Function Virtualization and Software Defined Networks, Berlin, Germany, 6-8 November 2017, pp. 1-7.

[20] B. Mirkhanzadeh et al, "An SDN-Enabled Multi-Layer Protection and Restoration Mechanism", Optical Switching and Networking, 2018

[21] Report, Open Edge Consortium, "OpenFog Reference Architecture for Fog Computing", https://www.openfogconsortium.org/ra/ (accessed: 2021)

[22] T. Zhang, "Fog Computing Brings New Business Opportunities and Disruptions", TechTarget, http://internetofthingsagenda.techtarget.com/blog/IoTAgenda/Fog-computing-brings-new-business-

opportunities-and-disruptions (accessed: 2021)

[23] F. Jalali et al., “Fog Computing May Help to Save Energy in Cloud Computing”, IEEE Journal on Selected Areas in Communications, Vol. 34, No. 5, 2016, pp. 1728-1739.

[24] S.-C. Hung, H. Hsu, S.-Y. Lien, K.-C. Chen, “Architecture Harmonization Between Cloud Radio Access Networks and Fog Networks”, IEEE Access, Vol. 3, 2015, pp. 3019-3034.

[25] A. Checko et al., “Cloud RAN for Mobile Networks - A Technology Overview”, IEEE communications surveys and tutorials, Vol. 17, No. 1, 2015, pp. 405-426.

[26] B. Lorenzo, J. Garcia-Rois, X. Li, J. Gonzalez-Castano, Y. Fang, “A Robust Dynamic Edge Network Architecture for the Internet of Things”. IEEE Network, Vol. 32, No. 1, 2018, pp. 8-15.

[27] D. Puthal et al., “Secure and Sustainable Load Balancing of Edge Data Centers in Fog Computing”, IEEE Communications Magazine, Vol. 56, No. 5, 2018, pp. 60-65.

[28] M. Syamkumar, P.Barford, and R. Durairajan, “Deployment Characteristics of The Edge in Mobile Edge Computing”, Proceedings of the 2018 Workshop on Mobile Edge Communications, August 2018, pp. 43-49.

[29] N. K. Giang, R. Lea, M. Blackstock, V. Leung, “Fog at the Edge: Experiences Building an Edge Computing Platform”, Proceedings of the IEEE International Conference on Edge Computing, San Francisco, CA, USA, 2-7 July 2018, pp. 1-9.

[30] O. Bibani, C. Mouradian, S. Yangui, R. H. Glitho, W. Gaaloul, N. B. Hadj-Alouane, M. Morrow, P. Polakos, “A demo of IoT healthcare application provisioning in hybrid cloud/fog environment”, Proceedings of the IEEE International Conference on Cloud Computing Technology and Science, Luxembourg, Luxembourg, 12-15 December 2016.

[31] CBS Interactive Report, "From Cloud to Edge: The next IT transformation", 2018, http://book.itep.ru/depository/fog_computing/SF_oct2018_edge.pdf (accessed: 2021)

[32] Reply Report, "From Cloud to Edge", 2020 https://www.reply.com/en/Shared%20Documents/from-cloud-to-edge-EN.pdf (accessed: 2021)

[33] W. F. Magalhães et. al., "Evaluating Edge-Cloud Computing Trade-Offs for Mobile Object Detection and Classification with Deep Learning", Journal of Information and Data Management, Vol. 11, No. 1, 2020.

[34] Technology Report, 2019, "The Next Step for Data Analytics - Driving Business Strategy", https://www.smartindustry.com/assets/Uploads/2019-Tech-Report-Data-Analytics.pdf (accessed: 2021)

[35] N. Meade, T. Islam, “Modelling and forecasting the diffusion of innovation – A 25-year review”, International Journal of Forecasting, Vol. 22, No. 3, 2006, pp. 519-545.

[36] J. S. Armstrong, “Selecting Forecasting Methods”, Principles of Forecasting: International Series in Operations Research & Management Science, Vol. 30, Springer, 2001.

[37] M. I. Sanchez, E. Zeydan, A. de la Oliva, A. S. Tan, U. Yabas, C. J. Bernardos, “Mobility management: Deployment and adaptability aspects through mobile data traffic analysis”, Computer Communications, Vol. 95, 2016, pp. 3-14.

[38] K.-W. Lim, S. Secci, L. Tabourier, B. Tebbani, “Characterizing and predicting mobile application usage”, Computer Communications, 2016

[39] Y. Leo, A. Busson, C. Sarraute, and E. Fleury, “Call detail records to characterize usages and mobility events of phone users,” Computer Communications, Vol. 95, 2016, pp. 82-94.

[40] M. J. Panik, “Growth Curve Modeling: Theory and Applications”, John Wiley & Sons, 2014.

[41] V. Križanović, “Telecommunications Services Selection Process Based on Analysis of Services Adoption”, Proceedings of the 15th Advanced International Conference on Telecommunications, France, Nice, France, 28 July – 1 August 2019, pp. 116-121.

[42] V. Križanović, D. Žagar, K. Grgić, and M. Vranješ, “Enhanced Predictive Modelling Process of Broadband Services Adoption Based on Time Series Data”, Advanced Engineering Informatics, Vol. 38, 2018, pp. 142-167.

An Optimized Distributed Routing Protocol for Energy Management Systems Based on Wireless Sensor Networks in Intelligent and Smart Structures

Original Scientific Paper

Ahmed Hammad
University of Mansoura,
Faculty of Engineering, Department of Electronics and Communications Engineering
Mansoura, Egypt
Khairtahmed@yahoo.com

M.A. Mohamed
University of Mansoura,
Faculty of Engineering, Department of Electronics and Communications Engineering
Mansoura, Egypt
mazim12@mans.edu.eg

Heba M. Abdel-Atty
University of Port Said,
Faculty of Engineering, Department of Electrical Engineering
Port Said, Egypt
Heba.Mohamed@eng.psu.edu.eg

Abstract *– Energy and spectrum efficiency for energy management systems based on wireless sensor networks in intelligent structures and powered by ambient energy harvesting (EH) are the main problems in wireless sensor networks. Herein, we consider relay selection methods. To address this issue, we proposed the optimal multiantenna power beacon opportunistic relay selection (OMPB-ORS) protocol, which uses decoding and forward methods, in which the relay wireless sensor nodes and the second source are energy-restricted and can harvest energy from a power beacon (PB) multiantenna to transmit aggregated information data from source to destination. The proposed protocol based on specific switching time receiver architecture enhances end-to-end performance for maximum hardware impairments and interference for the transceiver. To evaluate the performance, we compared our proposed protocol with best ORS (B-ORS), conventional ORS (C-ORS), and hybrid partial relay selection (H-PRS) protocols. Using the Rayleigh-fading channel, the simulation is driven based on asymptotic and exact form expressions of throughput (TP) and outage probability (OP). Simulation results show that the OMPB-ORS protocol achieves a higher TP and OP than all compared protocols.*

Keywords*: wireless sensor networks, energy harvesting, multiantenna power beacon, partial relay selection, opportunistic relay selection, hardware malfunctions.*

1. INTRODUCTION

Wireless sensor networks, which provide their energy from ambient energy harvesting (EH), have recently been listed as a promising technique to fix the famous problem of energy constraint for energy management systems in intelligent structures [1]. The communication equipment is outfitted with circuits that can harvest energy from the surrounding natural environment [1, 2]. In [3], the authors designed simultaneous wireless information and power transfer (SWIPT) systems. However, SWIPT systems are suited only for short-distance transmission because of a large operational sensitivity gap between the decoder and the energy harvester. In [4], the authors designed a novel system to deal with this issue in which power beacons (PBs) are used to activate wireless equipment. In [5, 6], to achieve maximum energy transfer and data rate for multiple input and output (MIMO) systems, the authors built a SWIPT receiver for the broadcasting system.

Recently, various PB-assisted wireless sensor networks using EH have been studied [7]. In [8], the authors proposed a novel hybrid wireless network with PBs deployed randomly in the area to offer mobiles an almost limitless battery life. In [7], using TDMA, the authors analyzed multiple-user wireless throughput (TP) for distributing Nakagami-m fading. Device-to-device (D2D) systems also deliberate PB-assisted techniques [9] because of the advantages of D2D systems, such as high spectral efficiency, low latency, and low-power transmission [10].

In contrast, besides the energy problem, the issue of spectrum scarcity needs a solution. In [11], the authors introduced the concept of cognitive radio (CR), where licensed primary users (PUs) can share their bands with unlicensed secondary users (SUs), provided the primary network's quality of service (QoS) is maintained. Generally, secondary users must know if PUs are available or not to use empty bands or shift to another spectrum [12, 13]. For CR WSNs, several spectrum-sensing models were created and compared [14, 15]. The benefits of CR WSNs and the significant differences between the three types of wireless sensor networks: CR WSNs, conventional WSNs, and ad hoc CR networks were also discussed in [16, 17]. Recently, various CR protocols were proposed and developed to ensure that SUs continue their operation [18, 19]. SUs are permitted to use the licensed bands simultaneously as PUs if the secondary transmitters adjust their broadcast power to meet a PU-imposed interference constraint. In [20,21], the authors improved the performance of the secondary network with vital technology, called cooperative relaying algorithms, due to the capacity to increase the performance gains. In [22], two relaying methods, such as partial relay selection (PRS) and opportunistic relay selection (ORS), have been extensively inspected. In PRS, the relay selection depends on the channel state information (CSI) for the source relay network. In ORS, the perfect relay must be selected to maximize the signal-to-noise ratio (SNR) end-to-end (e2e) between the transmitter and the receiver. In [23], the authors evaluated the performance of dual-hop CR WSNs in the existence of hardware noises and proposed three relaying algorithms, namely, best ORS (B-ORS), conventional ORS (C-ORS), and hybrid PRS (H-PRS) protocols. However, researchers proposed a better PRS scheme, where CSIs of the relay-destination connections are used to choose the relay [24]. In [23–29], several relay selection methods have been described in CR networks. Especially in [23], the PRS and ORS methods evaluated the overall performance in terms of bit error rate (BER) and outage probability (OP).

CR and EH were used in wireless sensor networks to solve the two main problems concerning energy and spectrum efficiency. In [30], the authors make SUs pick a channel to enter the harvested energy or data transmission. In [31], the authors use several SUs and various channels to fix the RF energy harvesting optimization problem for CR wireless networks. Specifically, a system model proposed by the authors where PUs take channels and make them busy, creating a chance for SUs to harvest energy and conserve it in the battery, then use the saved energy in the transmission process via an empty channel. In [32], the authors give a detailed analysis of the performance of a two-way CR EH-TWCR wireless network (EH-TWCR), which is based on decode-and-forward (DF) in the existence of transceiver limitations. In [19,20,24,25,33,34], the authors proposed the performance of multihop CR wireless networks, specifically end-to-end, where PB or RF signals of the primary transmitter can be used from SUs to harvest energy.

Because of the low-cost transceiver equipment, wireless sensor nodes fall victim to hardware limitations due to amplifier nonlinearities and phase noise [24, 26, 35, 36, 43]. The performance degradation can be recovered using a cooperative relaying algorithm. In [35], the effect of hardware failures on Nakagami-m fading channels in dual-hop relaying networks was investigated. In [36], in underlay CR networks, the performance of two-way relaying systems with hardware faults using EH relays were examined.

The contributions made by this work are as follows:

- A multiantenna PB wireless-powered cooperative communication network model is proposed, in which relays and sources are not connected with a fixed power network. Instead, PBs with multiantennas are used by relays and sources to harvest energy and then aggregated data to the destination.
- The proposed model is compared with three well-known relay selection models, namely, B-ORS, C-ORS, and H-PRS models, under interference and hardware limitations. The results compare the proposed optimal multiantenna power beacon ORS (OMPB-ORS) model performance with H-PRS, C-ORS, and B-ORS protocols in terms of system EH, TP, and OP.
- Numerical simulation is used to validate and drive outage the probability and average system TP for H-PRS, C-ORS, B-ORS, and proposed model closed-form expressions.
- The effects of multiantenna PB and other system characteristics, including harvesting time, number of relays, and position on system performance, are also examined.

The remainder of the paper is laid out as follows. The system model is described in section II. Next, Section III presents the conventional relay selection techniques. Furthermore, Section IV presents the equations of system performance for OMPB-ORS, H-PRS, C-ORS, and B-ORS schemes for OP and TP. Also, Section V displays the numerical result and compares the proposed models with all mentioned models. Finally, Section VI concludes the study.

2. SYSTEM MODEL

2.1 SYSTEM DESCRIPTION

Fig. 1. shows the proposed OMPB-ORS system model. It comprises the primary and secondary networks. The dual-hop technique communicates between a source S and a destination D in the primary network. P_n refers to the main licensed users, while n ∈ (1, 2,… , N). The secondary network comprises M relays, where N > 1 is denoted by Rm, while m ∈ (1, 2,… , M). Since the source has no direct connection with the destination, the system must select a suitable relay to transfer data from the source to the destination. Because they are considered to lack an integrated power supply, both and the set of M relay nodes must harvest energy from the multiantenna PB signal to allow information transmission. The source and relays have only one antenna. Additionally, they harvest energy from PBs. Two orthogonal time slots are used to transmit data via the chosen Relay.

It is assumed that Rayleigh fading affects all channels, and that the channel gains have exponential distributions. γSR_m and γDR_m are denoted as the channel gains for S ⟶ R_m and R_m ⟶ D links, respectively. $\gamma B_k R_m$ and $\gamma B_k S$ are denoted as channel gains between PB's k-th antenna and the S and relay R_m, respectively, where k = 1, 2, …, K denotes γSP_n and $\gamma R_m P_n$ as channel gains between S ⟶ P_n and R_m ⟶ P_n links. λ_{XY} is denoted as a random variable parameter, which equals $\lambda_{XY} = 1/E\{\gamma_{XY}\}$, where (X,Y) ∈ {S, R_m, B_k, P_n, D} and the anticipated value of the random variable Z is E{Z}.

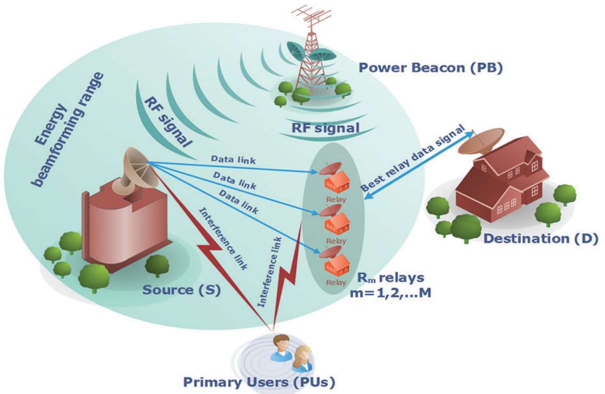


Fig. 1. System model of PB-assisted relaying protocols with relay selection methods.

The system model of the proposed protocol is implemented using the TS-HTC algorithm [37]. Fig. 2. shows that the protocol comprises three phases over the time block T, and only one node communicates at a time.

However, assuming optimal synchronization and channel state information in the network, it is beyond the scope herein to discuss how to achieve this synchronization. The batteries of S and R_m begin charging in the first phase, in which PB beamform RF signal to allow them to charge. Using the energy harvested in the first phase, S sends information to Rm. In the third phase, the best relay among the R_m relays is selected to transmit the received information from S to D by the proposed OMPB-ORS relay selection scheme or the well-known traditional relay selection schemes, i.e., H-PRS, C-ORS, or B-ORS as described in section 3.

The following set of assumptions is considered herein and in other related publications:

- A location-based clustering approach was used, in which the relays are clustered together close. This proposal is widely utilized in relay selection systems [38–40].
- As proposed in [41, 42], the PB is considered a network's devoted power source. The PB, S, R_m, and D nodes run in accordance with the harvesting energy and cooperating protocol.
- In the transmission phases, it is assumed that both the source and relay candidates exhaust their harvested energy.

2.2. HARDWARE MALFUNCTIONS

Assuming the transmitter (X) is connected with the receiver (Y), the signal-to-noise ratio of the X–Y connection can be obtained by (see [43]).

$$SNR_{XY} = \frac{P_X \gamma_{XY}}{(\tau_X^2 + \tau_Y^2) P_X \gamma_{XY} + N_0} = \frac{P_X \gamma_{XY}}{\tau_{XY}^2 P_X \gamma_{XY} + N_0} \quad (1)$$

where τ_X^2 and τ_Y^2 implement hardware malfunction levels at the transmitter and receiver, respectively, τ_{XY}^2 is the total hardware malfunctions level in the connection between transmitter and receiver, and N_0 is Gaussian noise variance at the receiver.

In the presence of Hardware malfunctions, the received signal of the X–Y link can be estimated as

$$y_{XY} = \sqrt{P_X} h_{XY} (s + \eta_{XY}) + \mu_{XY} + \upsilon_{XY} \quad (2)$$

where P_x is the power of transmitter X, h_{XY} is channel gain for X–Y link, μ_{XY} and η_{XY} are noises caused by hardware malfunctions in the receiver and transmitter, respectively, and ν_{XY} denotes the additive white Gaussian noises represented as Gaussian random variables with zero mean and variance N_0.

2.3. SIGNAL MODELING

2.3.1 EH phase

Here, the S and M relays charge their batteries by the beamform RF signal from PB, and the harvested energy by S and M relays can be formed, respectively, as

$$Q_s = \eta \alpha T P_B \sum_{k=1}^{K} \gamma B_k S \quad (3)$$

$$Q_{R_m} = \eta \alpha T P_B \sum_{k=1}^{K} \gamma B_k R_m \quad (4)$$

where PB represents the power of the transmitted signal from B, η is the efficiency of the harvested energy at S and M relays, and αT is the EH process time.

Fig. 2. shows that in the remaining (1 – α)T duration, the selected relay collaborates the source by decode and forwards the received signal. Finally, the optimal relay is chosen to transmit the S information to D once a relay selection procedure occurs. Consequently, the transmitted power at S and the set of R_m relays are expressed, respectively, as

$$E_s = \frac{Q_s}{2(1-\tau)/3} \tag{5}$$

$$E_{R_m} = \frac{Q_{R_m}}{(1-\tau)/3} \tag{6}$$

From [43] in the underlay CR with respect to interference constraint, the signal-to-noise ratio can be obtained at the 1st and 2nd hops across the relay provided by

$$SNR_{1m} = P_0 \gamma SR_m \big/ \tau_D^2 P_0 \gamma SR_m + N_0 \tag{7}$$

$$SNR_{2m} = P_m \gamma R_m D \big/ \tau_D^2 P_m \gamma R_m D + N_0 \tag{8}$$

where N_0 is the variance of the additive white Gaussian noise AWGN and

$$\Delta = P_B \big/ N_0 \tag{9}$$

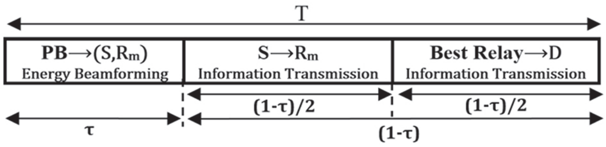


Fig. 2. Diagram of time-switching harvest then cooperate protocol (TS-HTC).

3. RELAY SELECTION SCHEMES

3.1 OPPORTUNISTIC RELAY SELECTION (ORS) SCHEME

Both channel hops are significant in the ORS relay selection method and should be considered [39, 41, and 42]. The optimal relay, which precisely maximizes the minimum number of channel strengths between S ⟶ R_m and R_m ⟶ D is selected and is provided by

$$R_S^{ORS} = \arg\max_{m \in M} \{\min(\gamma SR_m, \gamma R_m D)\} \tag{10}$$

3.2 PARTIAL RELAY SELECTION (PRS) SCHEME

This approach assumed that CSI is only valid for one hop [23, 39, and 44]. Precisely, when the CSI is available for the initial hop S ⟶ R_m, the PRS technique is denoted by PRSI. If the CSI is only accessible for the second chance R_m ⟶ D, it is termed PRSII. In PRSI and PRSII, the chosen relay can be represented as

$$R_S^{PRSI} = \arg\max_{m \in M} \{(\gamma SR_m)\} \tag{11}$$

$$R_S^{PRSII} = \arg\max_{m \in M} \{(\gamma SD)\} \tag{12}$$

3.3 PROPOSED OPTIMAL MULTIANTENNA POWER BEACON OPPORTUNISTIC RELAY SELECTION (OMPB-ORS) SCHEME

$$R_v : \min(SNR_{1v}, SNR_{2v}) = \max_{m=1,2,\ldots,M} (\min(SNR_{1m}, SNR_{2m})) \tag{13}$$

The optimal relay is chosen in the OMPB-ORS protocol to optimize the end-to-end SNR, i.e.,., where ν ∈ {1, 2,… , M}. The end-to-end performances of this scheme are then calculated as follows:

$$OP_{OMPB} = Pr(\,C_{th} > (1-\alpha)T\, log_2(\,1 + min(\,SNR_{1v}, SNR_{2v}))) \tag{14}$$

where the C_{th} in the secondary network is the desired data rate. The end-to-end channel capacity with decoding and forward technique of S ⟶ R_m ⟶ D path is described by

$$C_{SD} = (1-\alpha)T\, log_2(\,1 + min(\,SNR_{1m}, SNR_{2m})) \tag{15}$$

4. PERFORMANCE EVALUATION

4.1 OUTAGE PERFORMANCE OF PROPOSED SCHEME AND THROUGHPUT

The Flowchart of (OMPB-ORS) protocol scheme is shown in Fig. 3.The TS-HTC algorithm [37] was used in the proposed protocol, and the other three protocols used the TSR Protocol [23]. This approach made the difference in results clear, in favor of our protocol. The end-to-end OP can be defined as the probability that a positive threshold C_{th} exceeds the end-to-end capacity C_{SD}, and is expressed as follows:

$$OP_{OMPB} = \Pr(C_{th} > C_{SD}) \tag{16}$$

$$OP_{OMPB} = \Pr(\min(SNR_{1v}, SNR_{2v}) < \theta) \tag{17}$$

$$OP_{OMPB} = \Pr(\max_{m=1,2,\ldots,M} (\min(SNR_{1v}, SNR_{2v})) < \theta) \tag{18}$$

where,

$$\theta = 2^{\frac{2C_{th}}{(1-\alpha)T}} - 1 \tag{19}$$

Then, the throughput (TP) can be formulated as in [23]:

$$TP_{OMPB} = (1-\alpha)TC_{th}(1 - OP_{OMPB}) \tag{20}$$

where (1 – α)T is the overall transmission time from the source passed by the relay to the destination.

4.2 OUTAGE PERFORMANCE FOR (H-PRS), (C-ORS), AND (B-ORS) ALGORITHMS, AND THROUGHPUT

As in [23], the general form of e2e OP for the three protocols is given as

$$OP_U = 1 - \Pr(\min(SNR_{1m}, SNR_{2m}) \geq \theta) \qquad (21)$$

$$TP_U = (1-\alpha)TC_{th}(1-OP_U) \qquad (22)$$

where,

$$U \in \{H-PRS, B-ORS, C-ORS\} \qquad (23)$$

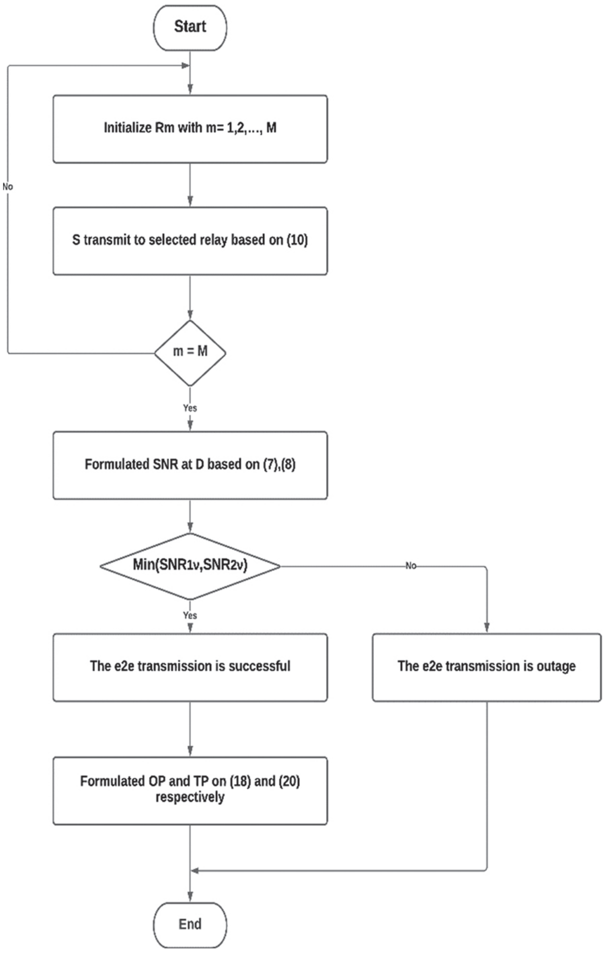

Start
Initialize Rm with m= 1,2,..., M
No
S transmit to selected relay based on (10)
m = M
Yes
Formulated SNR at D based on (7),(8)
Min(SNR1v,SNR2v)
No
Yes
The e2e transmission is successful
The e2e transmission is outage
Formulated OP and TP on (18) and (20) respectively
End

Fig. 3. The flow chart for the data transmission of OMPB-ORS scheme.

5. SIMULATION RESULTS

Here, the performance of the proposed protocol is presented. A set of numerical results is implemented under the existence of PUs provided with the interference constraints. To investigate the theoretical derivations, Monte-Carlo simulations are used. In TABLE 1, the WSN's nodes are organized in Cartesian coordinates in the simulation environment where S is located at the origin. The simulation, exact theoretical, and asymptotically theoretical results referred to them as (Sim), (Exact), and (Asym), respectively.

Table 1. System model parameters

System Parameters	Value
The number of relays	M = 2, 3, 4, and 5
The number of antennas of PB	K = 2
The transmission rate of S	C_{th}= 0.6, 0.7, and 1
Energy conversion efficiency	η = 1
Time block	T = 1
Harvesting time	α = 0.2s
Path-loss	β = 3
Ratio between Ith and PB	μ = 0.25
Number of PUs	N = 2
Relay coordinates	$(X_R,0)$
Destination coordinates	(1,0)
beacon coordinates	(0.5,0.5)
PU coordinates	(X_P, Y_P)

Fig. 4.compares the OP performance of the proposed protocol versus the H-PRS, C-ORS, and B-ORS protocols with C_{th} values. The proposed protocol has the lowest OP, and the H-PRS protocol has the highest. At a known high signal-to-noise ratio, the OP of the proposed, C-ORS and B-ORS protocols quickly decreased as Δ increased, which at Δ = 25 the enhancement percentages for the proposed protocol over H-PRS, C-ORS, and B-ORS protocols are 99.669%, 94.125%, and 94.20%, respectively, because the proposed, C-ORS and B-ORS protocols have a larger diversity gain than the H-PRS protocol.

To analyse the influence of distance on the proposed protocols' outage performance, OP was demonstrated as a function of the relay positions on the x-axis X_R. Fig. 5. Shows that the relays are in the best possible location, at which the proposed protocol OP value is lowest. Furthermore, when the relays are close to the destination, an intriguing consequence might be noticed, the OP values of the B-ORS and C-ORS protocols reach the OP of the proposed protocol. When the relays are extremely near the destination, the source-to-relay connection significantly impacts the OP of all protocols. Consequently, the B-ORS and C-ORS protocols are essentially equivalent to the proposed protocol.

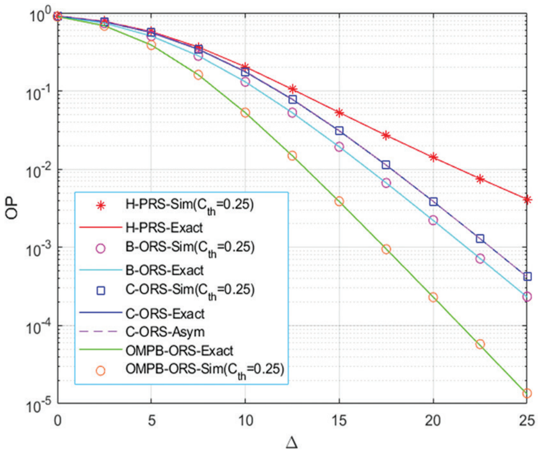


Fig. 4. OP as a function of Δ in dB when M = 2, X_R = 0.5, X_P = 0.5, Y_P = –0.5, α = 0.25, and $\tau_D{}^2 = \tau_1{}^2 = 0$

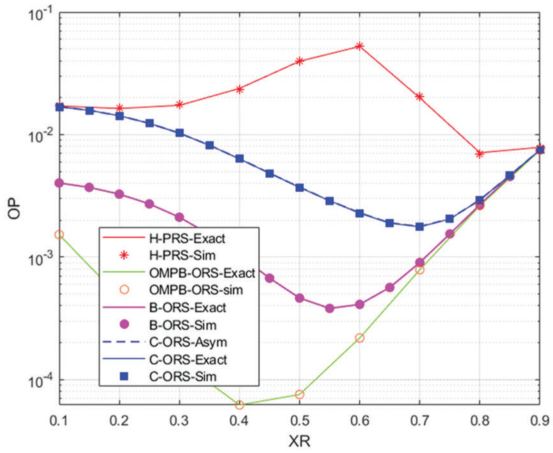


Fig. 5. OP as a function of X_R when $M = 4$, $X_P = 0.5$, $Y_P = -0.5$, $\alpha = 0.1$, $C_{th} = 0.6$, and $\tau_D^2 = 0.1$, and $\tau_1^2 = 0.05$.

Fig. 6. Explores the effect of the degree of hardware weakness τ_D^2 on the performance of all mentioned protocols. The OP values rapidly increase as τ_D^2 increases. Moreover, all protocols decrease when τ_D^2 exceeds 0.55.

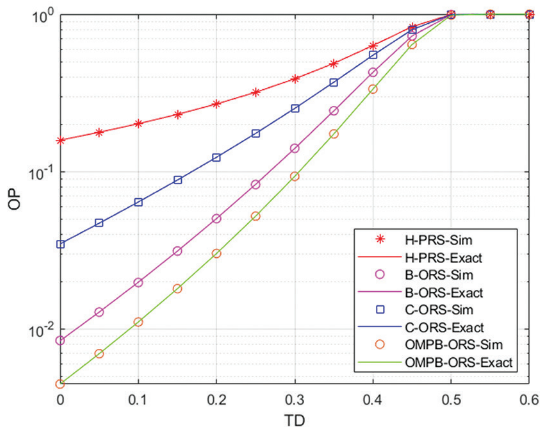


Fig. 6. OP as a function of τ_D^2 when $\Delta = 15$ dB, $M = 5$, $X_P = 0.5$, $Y_P = -0.5$, $\alpha = 0.1$, $C_{th} = 0.7$, and $\tau_1^2 = \tau_D^2 /2$.

Fig. 7. Shows that TP is plotted as a function of the time spent on the EH process. As previously said, α value acts as a significant function in the EH operation because it affects the collected and transmitted power of the source or chosen relay node. There exist optimum values of where the proposed protocol TP is the best (Fig. 7.). Consider the following example: when the α-value is extremely low, PB can only gather a limited amount of energy. Consequently, the source or relay node can only transmit information with a minimal quantity of energy. When the α-value is too high, the data are relayed from the source to the destination with a lower effective transmission time, which decreases the overall TP. Consequently, the best TP performance may be attained for practical design when an optimum α-value is obtained. Fig. 7. shows that the enhancement percentages at α = 0.035 for the proposed protocol over the H-PRS, C-ORS, and B-ORS protocols are 23.7%, 18.1%, and 8.3467%, respectively. Finally, similar to the OP measure, the proposed TP performance is always the highest overall values. Fig. 7. shows that the enhancement percentages at α = 0.035 for the proposed protocol over H-PRS, C-ORS, and B-ORS protocols are 3%, 6.6%, and 10.2%, respectively.

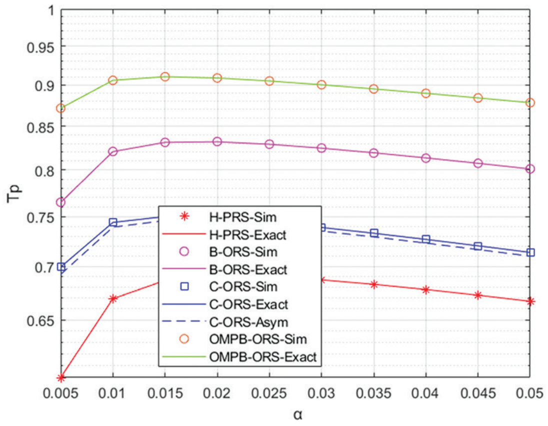


Fig. 7. TP as a function of α when $\Delta = 15$ dB, $M = 3$, $X_R = 0.5$, $X_P = 0.5$, $Y_P = -0.5$, $C_{th} = 1$, and $\tau_1^2 = \tau_D^2 = 0$

In Fig. 8., TP is shown against the number of relays. As predicted, increasing the M-value improves the TP of the OMPB-ORS, H-PRS, B-ORS, and C-ORS protocols. By effectively assigning the α-value, the performance of the investigated protocols can be enhanced.

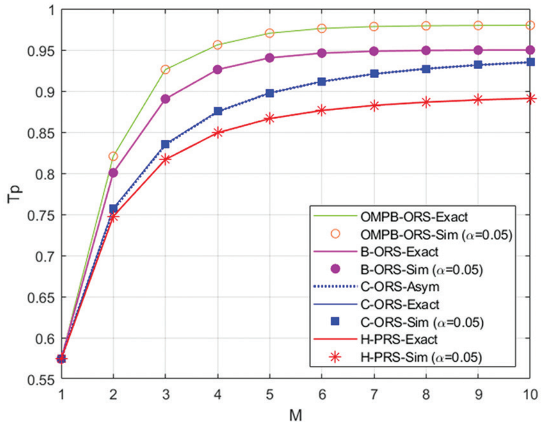


Fig. 8. TP as a function of M when $\Delta = 20$ dB, $M = 3$, $X_R = 0.4$, $X_P = 0.5$, $Y_P = -0.5$, $C_{th} = 1$, and $\tau_1^2 = 0.1$, $\tau_D^2 = 0.05$.

6. CONCLUSION

This study enhanced the performance of energy management systems based on WSN in intelligent structures under hardware weakness and interference restrictions. An OMPB-ORS protocol was proposed for EH relay networks using multiantenna PB, where PB supplies dual-hop DF relays and sources with RF signals to the EH process. In the presence of numerous PUs and across, i.e.,., Rayleigh-fading channels, exact and asymptotic formulations of the proposed protocol OP and TP were presented. The numerical results indicated that the OMPB-ORS protocol outperforms the B-ORS, C-ORS, and H-PRS protocols. Finally, by changing the energy harvesting ratio, increasing the number of relays, and locating the relays in the ideal place, the proposed protocol system's performance was improved.

7. REFERENCES

[1] R. L. Rosa, C. Dehollain, A. Burg, M. Costanza, P. Livreri, "An Energy-Autonomous Wireless Sensor with Simultaneous Energy Harvesting and Ambient Light Sensing", IEEE Sensors Journal, Vol. 21, No. 12, 2021, pp. 13744-13752.

[2] M. Li, C. Liu, Q. Li, "Energy Collaboration for Non-Homogeneous Energy Harvesting in Cooperative Wireless Sensor Networks", IEEE Access, Vol. 8, 2020, pp. 27027-27037.

[3] X. Liu, X. Yang, D. Ma, N. Jin, X. Lai, H. Tang, "A Novel Simultaneous Wireless Information and Power Transfer System", Proceedings of the IEEE Wireless Power Transfer Conference, 2019, pp. 212-215.

[4] C. Zhong, X. Chen, Z. Zhang, G. K. Karagiannidis, "Wireless-Powered Communications: Performance Analysis and Optimization", IEEE Transactions on Communications, Vol. 63, No. 12, 2015, pp. 5178-5190.

[5] C. Pan et al. "Intelligent Reflecting Surface Aided MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer", IEEE Journal on Selected Areas in Communications, Vol. 38, No. 8, 2020, pp. 1719-1734.

[6] R. Zhang, C. K. Ho, "MIMO broadcasting for simultaneous wireless information and power transfer", IEEE Transactions on Wireless Communications, Vol. 12, No. 5, 2013, pp. 1989-2001.

[7] N. P. Le, "Throughput Analysis of Power-Beacon-Assisted Energy Harvesting Wireless Systems Over Non-Identical Nakagami-Fading Channels", IEEE Communications Letters, Vol. 22, No. 4, 2018, pp. 840-843.

[8] K. Huang, V. K. N. Lau, "Enabling Wireless Power Transfer in Cellular Networks: Architecture, Modeling and Deployment", IEEE Transactions on Wireless Communications, Vol. 13, No. 2, 2014, pp. 902-912.

[9] Y. Liu, L. Wang, S. A. Raza Zaidi, M. Elkashlan, T. Q. Duong, "Secure D2D Communication in Large-Scale Cognitive Cellular Networks: A Wireless Power Transfer Model", IEEE Transactions on Communications, Vol. 64, No. 1, 2016, pp. 329-342.

[10] M. N. Tehrani, M. Uysal, H. Yanikomeroglu, "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions", IEEE Communications Magazine, Vol. 52, No. 5, 2014, pp. 86-92.

[11] E. O. Kandaurova, D. S. Chirov, "Intelligent Algorithms for Dynamic Spectrum Access a Secondary User in Cognitive Radio Systems", Proceeding of the Systems of Signal Synchronization, Generating and Processing in Telecommunications, 2021, pp. 1-6.

[12] F. Kong, J. Cho, B. Lee, "Optimizing Spectrum Sensing Time with Adaptive Sensing Interval for Energy-Efficient CRSNs", IEEE Sensors Journal, Vol. 17, No. 22, 2017, pp. 7578-7588.

[13] L. T. Dung, T. D. Hieu, S. G. Choi, B. S. Kim, "An B. Impact of Beamforming on the Path Connectivity in Cognitive Radio Ad Hoc Networks", Sensors, Vol. 17, No. 4, 2017, p. 690.

[14] X. Deng, P. Guan, C. Hei, F. Li, J. Liu, N. Xiong, "An Intelligent Resource Allocation Scheme in Energy Harvesting Cognitive Wireless Sensor Networks", IEEE Transactions on Network Science and Engineering, Vol. 8, No. 2, 2021, pp. 1900-1912.

[15] M. Zheng, C. Wang, M. Song, W. Liang, H. Yu, "SACR: A Stability-Aware Cluster-Based Routing Protocol for Cognitive Radio Sensor Networks", IEEE Sensors Journal, Vol. 21, No. 15, 2021, pp. 17350-17359.

[16] A. A. Khan, M. H. Rehmani, A. Rachedi, "Cognitive-Radio-Based Internet of Things: Applications, Architectures, Spectrum Related Functionalities, and Future Research Directions", IEEE Wireless Communications, Vol. 24, No. 3, 2017, pp. 17-25.

[17] A. Bagheri, A. Ebrahimzadeh, "Statistical Analysis of Lifetime in Wireless Cognitive Sensor Network for Multi-Channel Cooperative Spectrum Sensing", IEEE Sensors Journal, Vol. 21, No. 2, pp. 2412-2421, 15 Jan.15, 2021.

[18] N. T. Tung, P. M. Nam, P. T. Tin, "Performance evalution of two-way with energy harvestin and hardware noises", Digital Communications and Networks, Vol. 7, No. 1, pp. 45-54, 2020.

[19] M. N. Pham, “On the secrecy outage probability and performance trade-off of the multi-hop cognitive relay networks”, Telecommunication Systems, Vol. 73, 2020, pp. 349–358.

[20] S. Alvi, R. Hussain, Q. Hasan, S. Malik, “Improved buffer-aided multi-hop relaying with reduced outage and packet delay in cognitive radio networks”, Electronics, Vol. 8, No. 8, 2019, p. 89.

[21] Shakeel A. Alvi, Riaz Hussain, Atif Shakeel, Muhammad Awais Javed, Qadeer Ul Hasan, Byung Moo Lee, Shahzad A. Malik, “QoS-Oriented Optimal Relay Selection in Cognitive Radio Networks”, Wireless Communications and Mobile Computing, Vol. 2021, Article ID 5580963, 15 pages, 2021.

[22] D. H. Ha, T. N. Nguyen, M. H. Q. Tran, X. Li, P. T. Tran, M. Voznak, “Security and Reliability Analysis of a Two-Way Half-Duplex Wireless Relaying Network Using Partial Relay Selection and Hybrid TPSR Energy Harvesting at Relay Nodes”, IEEE Access, Vol. 8, 2020, pp. 187165-187181.

[23] T. D. Hieu, T. T. Duy, L. T. Dung, S. G. Choi 2018. "Performance Evaluation of Relay Selection Schemes in Beacon-Assisted Dual-Hop Cognitive Radio Wireless Sensor Networks under Impact of Hardware Noises", Sensors, Vol. 18, No. 6, p. 1843.

[24] T. T. Duy, P. T. D. Ngoc, P. T. Tran, “Performance Enhancement for Multihop Cognitive DF and AF Relaying Protocols under Joint Impact of Interference and Hardware Noises: NOMA for Primary Network and Best-Path Selection for Secondary Network”, Wireless Communications and Mobile Computing, Vol. 2021, Article ID 8861725, 15 pages, 2021.

[25] P. M. Quang, T. T. Duy, V. N. Q. Bao, “Performance Evaluation of Radio Frequency Energy Harvesting-Aided Multi-hop Cooperative Transmission Networks”, Proceedings of the 25th Asia-Pacific Conference on Communications, 2019, pp. 521-526.

[26] H. V. Toan, Hoang, T. M. Duy, L. T. Dung, “Outage Probability and Ergodic Capacity of a Two-User NOMA Relaying System with an Energy Harvesting Full-Duplex Relay and Its Interference at the Near User”, Sensors, Vol. 20, 2020, p. 6472.

[27] P. N. Son, T. T. Duy, K. Ho-Van, “SIC-Coding Schemes for Underlay Two-Way Relaying Cognitive Networks”, Wireless Communications and Mobile Computing, Vol. 2020, Article ID 8860551, 17 pages, 2020.

[28] P. N. Son, T. T. Duy, “A new approach for two-way relaying networks: improving performance by successive interference cancellation, digital network coding and opportunistic relay selection”, Wireless Networks, Vol. 26, 2020, pp. 1315–1329.

[29] T. N. Nguyen, T. H. Minh, P. T. Tran, M. Voznak, T. T. Duy, T. L. Nguyen, P. T. Tin, “Performance Enhancement for Energy Harvesting Based Two-way Relay Protocols in Wireless Ad-hoc Networks with Partial and Full Relay Selection Methods”, Ad Hoc Networks, Vol. 84, 2019, pp. 178–187.

[30] D. T. Hoang, D. Niyato, P. Wang, D. I. Kim, “Opportunistic channel access and RF energy harvesting in cognitive radio networks”, IEEE Journal on Selected Areas in Communication, Vol. 32, 2014, pp. 2039–2052.

[31] D. T. Hoang, D. Niyato, P. Wang, D. I. Kim, “Performance analysis of wireless energy harvesting cognitive radio networks under smart jamming attacks”, IEEE Transactions on Cognitive Communication Networks, Vol. 1, 2015, pp. 200–216.

[32] D. K. Nguyen, D. Jayakody, S. Chatzinotas, J. S. Thompson, J. Li, “Wireless energy harvesting assisted two-way cognitive relay networks: Protocol design and performance analysis”, IEEE Access, Vol. 5, 2017, pp. 21447–21460.

[33] C. Xu, M. Zheng, W. Liang, H. Yu, Y. C. Liang, “Outage performance of underlay multihop cognitive relay networks with energy harvesting”, IEEE Communication Letters, Vol. 20, 2016, pp. 1148–1151.

[34] C. Xu, M. Zheng, W. Liang, H. Yu, Y. C. Liang, “End-to-end throughput maximization for underlay multi-hop cognitive radio networks with RF energy harvesting”, IEEE Transactions on Wireless Communications, Vol. 16, 2017, pp. 3561–3572.

[35] E. Bjornson, M. Matthaiou, M. A. Debbah, “A new look at dual-hop relaying: Performance limits with hardware impairments”, IEEE Transactions on Communication, Vol. 61, 2013, pp. 4512–4525.

[36] O. Messadi, A. Sali, V. Khodamoradi, A. Salah, G. Pan, S. J. Hashim, N. K. Noordin, “Optimal Relay Selection Scheme with Multiantenna Power Beacon

for Wireless-Powered Cooperation Communication Networks", Sensors, Vol. 21, 2021, p. 147.

[37] T. Yuan, M. Liu, Y. Feng, "Performance Analysis for SWIPT Cooperative DF Communication Systems with Hybrid Receiver and Non-Linear Energy Harvesting Model", Sensors, Vol. 20, 2020, p. 2472.

[38] H. Chen, Y. Li, J. L. Rebelatto, B. F. Uchoa-Filho,; B. Vucetic, "Harvest-then-cooperate: Wireless-powered cooperative communications", IEEE Transactions on Signal Processing, Vol. 63, 2015, pp. 1700–1711.

[39] T. N. Nguyen, M. Tran, D. H. Ha, T. T. Trang, M. Voznák, "Multi-source in DF Cooperative Networks with the PSR Protocol Based Full-Duplex Energy Harvesting over a Rayleigh Fading Channel: Performance Analysis", Proceedings of the Estonian Academy of Sciences, Vol. 68, 2019, pp. 264–275.

[40] P. T. Tin, B. H. Dinh, T. N. Nguyen, D. H. Ha, T. T. Trang, "Power Beacon-Assisted Energy Harvesting Wireless Physical Layer Cooperative Relaying Networks: Performance Analysis", Symmetry, Vol. 12, 2020, p. 106.

[41] T. M. Hoang, N. T. Tan, X. N. Tran, "Performance analysis of power beacon-assisted energy harvesting NOMA multi-user relaying system over Nakagami-m fading channels", International Journal of Electronics and Communications, Vol. 115, 2020, p. 153022.

[42] P. K. Sharma, P. K. Upadhyay, "Cognitive relaying with transceiver hardware impairments under interference constraints", IEEE Communication Letters, Vol. 20, 2016, pp. 820–823.

[43] N. T. Do, V. Bao, B. An, "Outage performance analysis of relay selection schemes in wireless energy harvesting cooperative networks over non-identical rayleigh fading channels", Sensors, Vol. 16, 2016, p. 295.

Development and Hardware Implementation of IoT-Based Patrol Robot for Remote Gas Leak Inspection

Original Scientific Paper

Tasneem Yousif
Computer Engineering Department Laboratories
Department of Computer Engineering, University of Bahrain
Skhair, Kingdom of Bahrain
tasneemyousif@hotmail.com

Wael El-Medany
Computer Engineering Department Laboratories
Department of Computer Engineering, University of Bahrain
Skhair, Kingdom of Bahrain
welmedany@uob.edu.bh

Abstract *– The Internet of Things Robot (IoTR) is an emerging paradigm that brings together robotic systems with the Internet of Things (IoT) that connect sensors and smart objects pervasively embedded in everyday environments. With the recent developments in robotic system applications, it becomes apparent that the mobile robot has great importance in real-world applications such as navigation and surveillance. One of the most important applications of a mobile robot is patrolling and gas leak detection. This paper proposes a real-time IoT Robot (IoTR) that can be used indoors or outdoors for gas leak detection purposes. The proposed mobile robot is equipped with microphones, speakers, the hub of smart sensors that are necessary for patrolling and gas leak detection, a high-resolution IP video camera for live video streaming, Bluetooth for indoor applications and tracking, and GPS/GPRS for outdoor applications and tracking. The experimental testing of the preliminary prototype confirms the design objectives. The robot has been tested for indoor and outdoor modes; the robot can detect gas leakage and provides a live video streaming of the surrounding area, which can be tracked on Google maps. At the same time, the robot can be controlled remotely through a mobile app or website, the robot can move autonomously and avoid obstacles. The proposed work provides a low-cost IoT robot through the use of the available and cheap components and sensors, which featured a high quality at the same time. Our proposed system exhibits promising gas sensing performance in harsh environments, using intelligent gas sensors that have a fast response (>10s), low cost, high sensitivity, long life, robustness, and physical size.*

Keywords*: Gas Leak Detection, IoT Robot, Remote Monitoring, Patrol Robot, Tracking System, Real-Time*

1. INTRODUCTION

With the development of society, economy and rapid developments of autonomous mobile robots that have been developed to tackle the challenges of the petroleum industry, increasing the protection measurements, enhancing the quality of critical patrol and gas areas and decreasing costs are now possible. The use of security patrol robots has become very important recently because it helps to prevent human lives from danger and reduce human errors. The patrol robot can be used indoors or outdoors depending on the applications, with automatic obstacle avoidance as a common feature [1, 2]. Oil and petroleum companies need a safe environment for their critical work. The industry of oil and gas is now positively looking for advanced robotic solutions in conjunction with the growth of global demand and depleting resources for fossil fuels. These smart robotic systems are used to increase their productivity and safety [3]. Health and safety awareness is so essential for all workers and technicians who are working in those companies [4, 5].

Explosive and toxic gases are surrounding us everywhere such as gas stations, power plants, landfill sites, hotels, kitchens, wells, oil, and gas companies [3]. They have different sources such as welding, swamp wells, volcanos, grinding, mining, petroleum areas, etc. Explosive materials, toxic wastes, and hazardous gases endanger our lives. They cause chronic and dangerous diseases for humans such as pneumonia and angina pectoris. Also, it may cause instant death for humans. Moreover, accidents that happen due to these hazardous gases cause property damage, substantial money, injuries, grieving families, and many other fatalities [3, 6, 7]. These gases are very hazardous for companies in the oil and gas industry due to their critical job. Therefore, oil and gas companies need to comply with strict gas safety regulations to ensure no gas leakage occurs. The employees and the people working in these companies deserve to work in a safe and healthy environment. Our solution is a creative cost-effective and scalable solution for health and safety environments to employees and employers on-site, potentially from remote sites.

The presented work in this paper is extended research of the proposed remote monitoring system that has been introduced by [8, 9]. The research aims to develop real-time [10] remote monitoring and control IoT robot that is used for security and detecting explosive gases at the same time. The robot is equipped with a video camera and smart sensors for the smart detection tasks carried out by the robot. The proposed system consists of different subsystems, the controlling unit where it is either indoors using Bluetooth connection or outdoor controlling using the internet connection, the monitoring unit with a live streaming video camera, and the detection unit that consists of smart sensors and is used for detecting explosive gases. The IoT robot will work as security surveillance in real-time and can be tracked and controlled remotely through the Global Positioning System (GPS) and General Packet Radio Service (GPRS) network [11]. The robot will be able to change its route once the stationary alarm sensors are triggered. This feature will help the pipeline inspectors and the patrolling employees to maintain a safe distance in the remote areas of the site in case of any hazardous gas leakage. Additionally, the robotic system gives instant data analysis of the surrounding gases for the users using data visualization dashboards.

The paper is organized as follows: Section two addresses the state-of-the-art, where related work is described. Section three outlines the existing problem and the corresponding solution provided by the IoT robot, by describing the full system architecture and the main subsystems. Section four outlines the software user interface. In section five, the testing and validation results are discussed. Finally, section six provides conclusions and future research.

2. RELATED WORK

Gas leakage detection in industrial facilities is critical for ensuring the safety of human life, stationary gas sensors can be used to discover the gas leaks. In the following section, a review of the state-of-the-art mobile robot applications in security and gas leakage will be covered. A mobile robotic system for remote leak sensing and localization in large industrial environments has been presented by [12] to develop the RoboGas Inspector shown in Fig. 1. The gas leak detection technology for LNG facilities in Australia is shown in Fig. 2.

In [13], the authors introduced a novel smart security robot that used to work in dark non-vision environments, since most of the methods available are not that capable to work in a dark environment. The proposed work uses fuzzy logic and neural network methods to determine the situations that are not normal in the environment area, and they used a different level of danger alarm. The research shows a general view of the proposed security system as shown in Fig. 3. The system starts with the initialization step (all devices will be turned on), then the path planning operation, and then a motion control operation using four different fuzzy logic modules to monitor the area. During this stage, if any abnormal thing happens, the system will be switched to the alarm system to alarm the user.

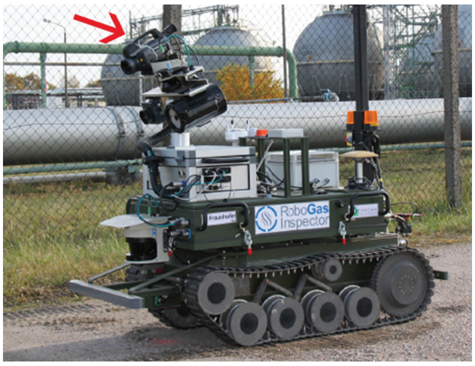


Fig. 1. The RoboGas Inspector [12]

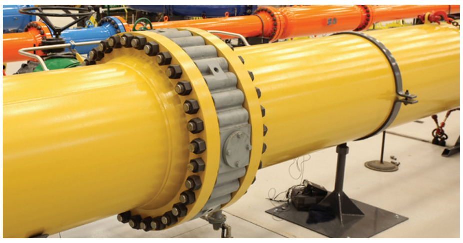


Fig. 2. Gas-Leak Detection Technology for LNG Facilities in Australia [12]

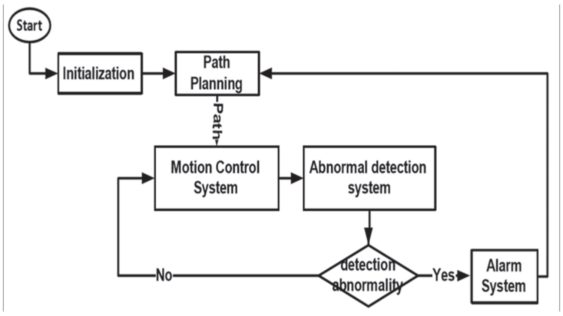


Fig. 3. General view Proposed Security System [13]

The authors in [14] proposed an android controlled surveillance spy robot with a wireless night vision camera. They aim to save human lives in the military sector where this robot is used in wars to monitor the area where it is dangerous for the army. Their robot consists of a night vision camera fixed to it to monitor the area while saving human lives. This robot is an Arduino-based controlled robot that used a developed android application for controlling it. However, their system does not have a gas detection system.

A fire robot based on a quad-copter has been developed in [15]. This robot is required to detect fires. They used an algorithm for processing the video signal to detect the fire and for the robot realization. Also, they used a software program using the SDK software developer kit to control the robot's motion.

The research introduced by [16] aims to develop a surveillance system that uses an unmanned aerial vehicle (UAV). The UAV robot has mobility capability. The authors proposed a surveillance robot for indoor monitoring. They combined six components of the system, such as unmanned aerial vehicles, vision-based pose estimation, vision-based state estimation, patrol path planning, UAV controller, joystick controller, and priority assignment for different inputs of the robot[16]. These components are used for autonomous patrol and surveillance systems. They developed and tested their system in the indoor environment. Fig. 4 shows the system architecture of the system.

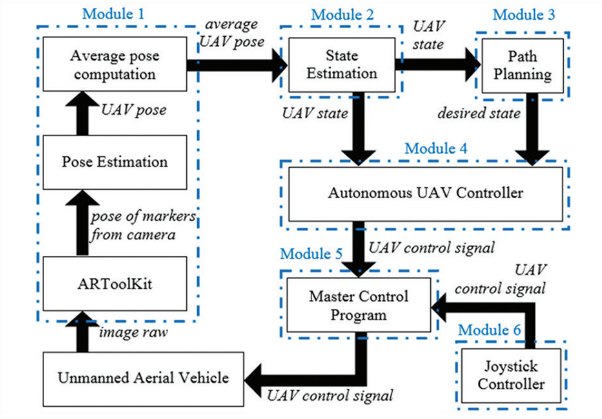


Fig. 4. System Architecture of the autonomous patrol and surveillance system using UAV [16]

In [17], the system is integrated with a mobile robot that can detect carbon monoxide (CO) with the temperature. The authors stated that their system tackled the challenges of the current CO gas sensors which are the limited detection ranges and sensitivity due to environmental factors such as temperature and humidity. This system lacks flexibility due to the usage of one gas sensor compared to our system. Besides, it does not have outdoor applications which are fundamental to industrial and landfill applications.

The authors in [18] introduced a mobile robot with a gas leakage detector for safety purposes. The system has an array of 16-metal oxide sensors (e-nose) that can detect the concertation of the gases by obtaining 16 voltage measurements. The system focused on detecting gas leakage in closed rooms due to the few amounts of airflow surrounding the rooms.

The authors proposed natural gas detection by using an intelligent gas searching method achieved by a swarm optimization algorithm [19]. Their system has a detection strategy for gas detection, gas tracking, and gas source localization with search time consumed between 27s-92s. The drawback of the discussed research is that the detection system has not yet been established with a real experimental robot based on relevant hardware and software equipment. Furthermore, their system needs more optimization for the user searching algorithm and a real simulation platform for the gas sensors analysis.

The system proposed in [20], provided a fireNose that contains three Metal-Oxide (MOX) sensors (SnO2, WO3, and NiO). They used online unsupervised gas discrimination algorithms that increased the sensitivity for the gas sensors.

2.1 COMPARISON TO STATE-OF-THE-ART RELATED WORK

The proposed design has been compared to state-of-the-art related work; the comparison has been carried out with other systems that are similar to the proposed system. The gas inspector system that has been presented by [21] and shown in Fig. 1 uses multiple measurement principles for remote sensing. This solution is used for the detection of hazardous gases by using novel leak-detection technologies. However, this system is portable which will make human life endanger. The explosive detector introduced by [22], is using Fido® XT. The Fido solution uses polymer-based technology to achieve faster detection results. However, this solution is a handheld detector. The IoTR has more advantages than stationary sensors network (SN) and portable sensors. The IoTR has the flexibility and feasibility to combine a variety of gas sensors depending on the application. Besides, the IoTR has the feature to explore the environment and detect gas leakages remotely. In contrast, handheld sensors have to be carried by the users to the required areas to detect the hazard gases, thus this jeopardizes human lives. The IoTR will be equipped with intelligent sensors to provide accurate computational gas results, gas sources localization, live tracking system, real-time visual monitoring, map-based approaches, alarming system, and remote-controlling. On the other hand, the portable sensors and stationary sensors have a limitation of inability to sense accurately when there is a variance in optimal sensor positions across long distances. Our system is a real-time system with a tracking system, live video streaming, and a web-based or mobile application to access the sensors reading remotely with intelligent dashboard visualizations. Furthermore, our mobile robot is operated remotely from anywhere on the planet, therefore, our system emphasizes the outdoor controlling. On the other hand, other solutions use limited distance controlling. Additionally, the gas sensors that have been used in the proposed system have advantages of low cost, fast time response, which is less than 10s, stability, physical size, and long life. Moreover, our robotic system has a reasonable cost compared to other solutions in the market. Our IoTR cost is ≈1000$.

3. IoTR ARCHITECTURE

The proposed IoT robot shown in Fig. 5 handles different kinds of tasks depending on the usage and the application. The proposed mobile robot has the flexibility to be equipped with a specific sensor based on the application. The robot can be remotely controlled with the mobile application or through the website to go through narrow or tough locations to discover the gas leakages in its surrounding area. It is equipped with a tracking device and a live video streaming camera for real-time monitoring. Additionally, the robotic system provides instant notifications through text messages on cell phones or text emails. Also, the system provides a voice alarming once the sensors trigger high concertation of the hazard gases. Besides, the robotic system provides gas sensing analysis reports on a daily, monthly, and yearly basis using visualization dashboards [23, 24]. The robotic system architecture for relaying data between the robot and the monitoring server is illustrated in Fig. 6. The patrol robot system depends on the IoT and cloud computing features. The system has database storage on the cloud server which will store all the data required for the system. These data contain the gas sensing results, the temperature measurements, the humidity measurements, the GPS coordinates data, the response time, the latest controlling request that is required for remote controlling, and the live streaming video. This full safety system sends and receives requests through GPRS/GSM communications. Fig. 7 shows the main hardware components required for real-time monitoring, tracking, and controlling, including Arduino Microcontroller (Fig. 7 (a)), Bluetooth shield (Fig. 7 (b)), GPRS shield (Fig. 7 (c)) and GPS shield (Fig. 7 (d)), all interfaced with the IoT robot. Fig. 8, shows the State Machine Chart (SMC) for the microcontroller Interfaced with the complete System. SMC illustrates the two different ways of accessing the system which are accessing the mobile application and websites. The mobile application has the function of indoor controlling using Bluetooth and autonomous controlling using a ready path follower. The website has three main functions: outdoor controlling, live tracking, and gas sensing results. All of these functions were successfully achieved through the communication between the GPRS and the web-server. Besides, there is a function of live video streaming using an IP camera to be accessed through the website or mobile application.

The patrol robot consists of four main subsystems:

1. Live video streaming using IP camera
2. Real-time remote controlling system whether indoor or outdoor.
3. Real-time tracking using GPS/GPRS and Google Map technologies.
4. Gas leakage inspection system using intelligent gas sensors.

The main four subsystems are discussed in the following subsections.

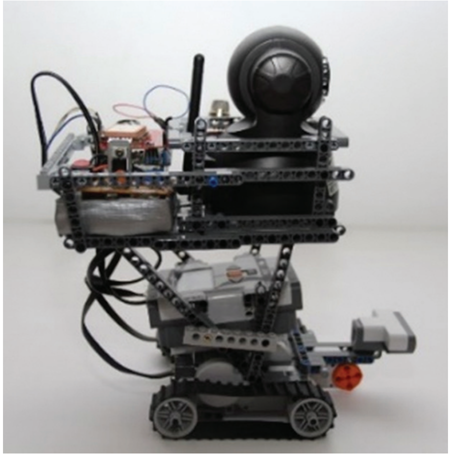


Fig. 5. IoT Robot Prototype

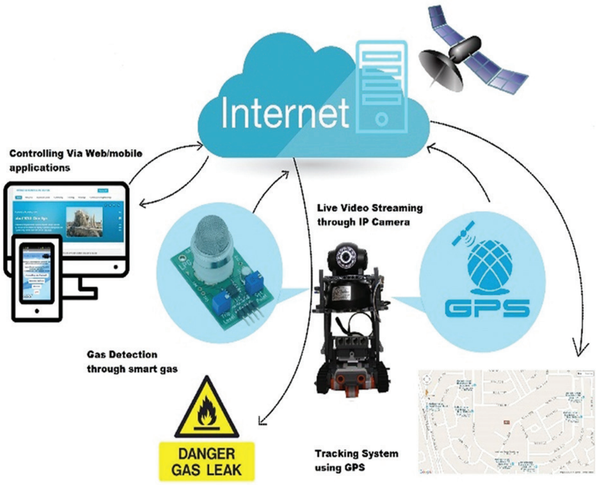
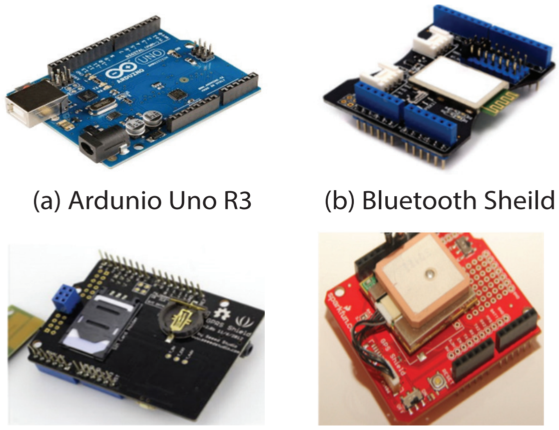

Internet
Controlling Via Web/mobile applications
Live Video Streaming through IP Camera
GPS
Gas Detection through smart gas
DANGER GAS LEAK
Tracking System using GPS

Fig. 6. IoTR System Architecture



(a) Ardunio Uno R3 (b) Bluetooth Sheild

(c) GPRS Sheild (d) GPS Sheild

Fig. 7. Hardware components interfaced with the IoTR

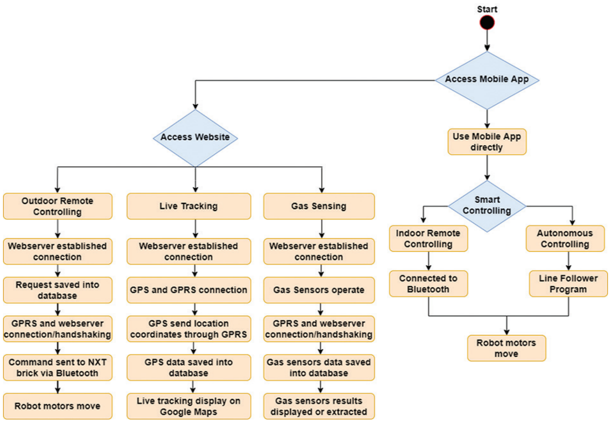

Start
Access Mobile App
Access Website
Use Mobile App directly
Outdoor Remote Controlling
Live Tracking
Gas Sensing
Smart Controlling
Indoor Remote Controlling
Autonomous Controlling
Webserver established connection
Webserver established connection
Webserver established connection
Connected to Bluetooth
Line Follower Program
Request saved into database
GPS and GPRS connection
Gas Sensors operate
Robot motors move
GPRS and webserver connection/handshaking
GPS send location coordinates through GPRS
GPRS and webserver connection/handshaking
Command sent to NXT brick via Bluetooth
GPS data saved into database
Gas sensors data saved into database
Robot motors move
Live tracking display on Google Maps
Gas sensors results displayed or extracted

Fig. 8. State Machine Chart for the microcontroller Interfaced with the complete System

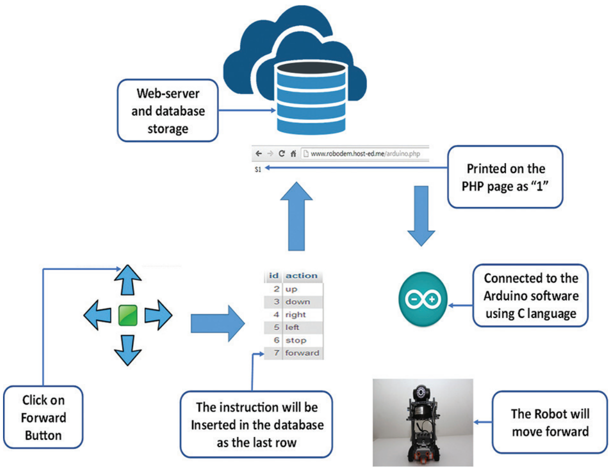
3.1 REAL-TIME VISUAL MONITORING

The patrol robot has one eye to provide a real-time visual monitoring of the robot's location and the surrounding area. The IP camera shown in Fig. 9 has been used for this purpose. It has several features such as, it can be accessed through the internet, has night visibility up to 15 or 20 meters, a low-cost, support smartphone, it is equipped with a motion sensor for automatic triggering of the camera, and can take a snapshot or record live video streaming. This robotic system can be used remotely from homes or work areas. Therefore, the camera has been accessed remotely through a static IP address from an Internet Service Provider (ISP) which is required for the stability and security of the network. As the main purpose of this robotic system is security, therefore, one user can access the camera per account. A DNS (Domain Name System) has been assigned to the IP address. The user account is synchronized with the IP address by registering that account in the router with the specific port address to allow any incoming connection to access the IP address through that port.

Fig. 9. IP Camera

3.2 REAL-TIME REMOTE CONTROLLING

To control the movement of the robot, some commands must be sent from the mobile apps or web-server to move it forward, backward, right, left, or stop its movement [25, 26]. These commands will be saved in the web-server's database. The robot moves in a specific direction through a series of communications and handshaking that have been done between the web-server, GPRS, and Bluetooth modules.


Web-server and database storage
www.robodem.host-ed.me/arduino.php
$1
Printed on the PHP page as "1"
Connected to the Arduino software using C language
Click on Forward Button
The instruction will be Inserted in the database as the last row
The Robot will move forward

id	action
2	up
3	down
4	right
5	left
6	stop
7	forward

Fig. 10. Outdoor Controlling Process Steps

As the main purpose of this robotic system is to sense dangerous areas, outdoor controlling has been accomplished to save human lives. The robot has been controlled remotely by a developed web application by giving instructions for spatial movement or stopping. These instructions are represented by specific commands that are understandable by the NXT intelligent brick of the robot. As illustrated in Fig. 10, the user will click on one of the buttons to control the robot by moving it forward, backward, left, right, or stop it. Each request is associated with that clicked button will be received by the web-server. Then, the web-server network will send the request to be saved in the web-server database storage. The instruction will be saved in the last row at the instructions table. After that, the handshaking will be accrued between the web server and the GPRS. Then, the web-server will send the command through a generated PHP code that contains the movement command with a unique code. For instance, forward instruction is represented by '1' code. Each code has its response to be sent through the HTTP protocol between the web-server, the connected GPRS, Bluetooth, and Arduino. After that, the GPRS will send the response to the Bluetooth module. Finally, the motors of the robot will be able to move or stop based on each command. Fig. 11 shows the flowchart of the controlling process.

The goal of outdoor controlling cannot be achieved through the GPRS shield only because LEGO Mindstorm NXT robotics cannot be connected to the GPRS directly. LEGO Mindstorm NXT has only the ability to communicate with other devices and shields via Bluetooth. For this reason, a Bluetooth shield is used to integrate a Bluetooth module that can receive the data from the GPRS and send it to the robot by connecting the Bluetooth shield with the GPRS shield. Fig. 12 shows the hardware components required for real-time monitoring and controlling. The general architecture of the Bluetooth protocol telegram is shown in Fig. 13. Byte 0 contains the telegram type. The three types of telegrams (byte 1, 2, 3) represent the direct command telegram, reply command telegram, and the system command telegram respectively.

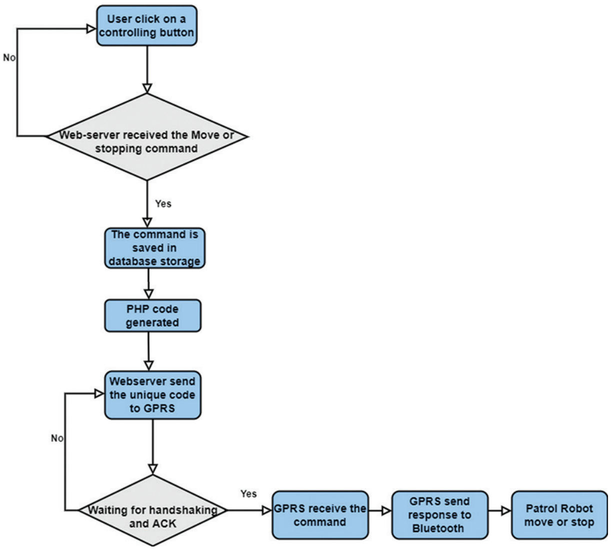

User click on a controlling button
No
Web-server received the Move or stopping command
Yes
The command is saved in database storage
PHP code generated
Webserver send the unique code to GPRS
No
Waiting for handshaking and ACK
Yes
GPRS receive the command
GPRS send response to Bluetooth
Patrol Robot move or stop

Fig. 11. Flowchart of the outdoor controlling process

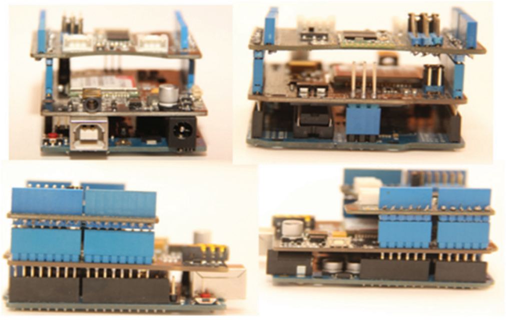


Fig. 12. Connected GPRS shield, Bluetooth shield, and Arduino UNO

For this system design, the direct command and the reply command have been used. Table 1 shows some important commands with their types that have been used for the robotic system. The other bytes consist of the command itself and the reply command based on each telegram type. The direct commands have a limited size which is 64 bytes including the byte that represents the telegram type. Two additional bytes are not included in the size limit, and they should be in front of the Bluetooth message as shown in Fig. 14. Based on testing and experiments, the motor's medium speed was set to 360 degrees/seconds. Fig. 15 shows that the patrol robot has reached a distance of 100m which is represented by the objective (green mark) within 10.2s. Therefore, the speed is accelerated by 9.79 m/s. While the patrol robot reached the same objective (red mark) within 5.1s with a fast speed of 19.58 m/s. In this system design, two motors have been connected to ports B and C of the NXT Brick. There are lists of commands used for the spatial movement of the patrol robot are listed below [27]:

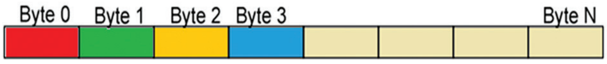
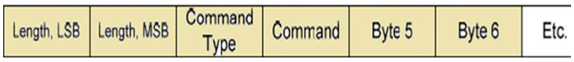


Fig. 13. The General Architecture of the Bluetooth Protocol of LEGO Mindstorm NXT



Length, LSB	Length, MSB	Command Type	Command	Byte 5	Byte 6	Etc.

Fig. 14. Bluetooth Message Architecture

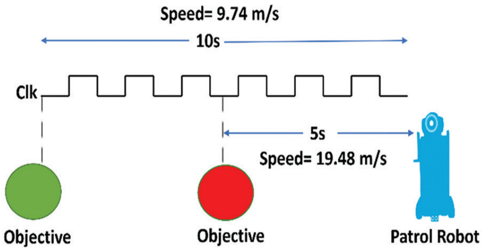


Fig. 15. Speed test bench for the IoTR movement

I. Moving the robot forward with fast speed

The command shown in Fig. 16-a is used for moving the robot forward with fast speed. The command consists of 14 bytes. The first two bytes contain the length of the message without counting these two bytes, the command will contain 12 bytes in decimal. The number should be converted to hexadecimal to be 0x0C in the first byte. The second byte contains 0x00 because one byte is enough for representing 12 in hexadecimal. The third byte has the value of 0x80 which is the command type that is used to send direct commands without waiting for a response. The fourth byte contains 0xff which means moving all motors. The fifth byte contains 0x64 in hexadecimal which is the turn ratio, and it has a range (-100 to 100) in decimal. This value should be positive to move the robot forward. The maximum speed here is 100 in decimal (64 in hexadecimal). The sixth byte contains 0x07, which are used for the mode byte which means turn on all motors. The remaining bytes will be the same for all commands. The meaning for 0x00 in the seventh byte is disabling the regulation. In the eighth byte, it is for clearing the turn ratio to move straight. The ninth byte has 0x20 enumerations which are used for setting the output to be run. Finally, the last four bytes which contain (0x00 0x00 0x00 0x00) are used for the taco limit to continue running indefinitely.

II. Move the robot forward with medium speed

This command moves the robot forward with medium speed, Fig. 16-b shows the command. This command is similar to the command of (moving the robot forward with fast speed) with one difference in the fifth byte. The range is (-100 to 100) in decimal, the half of this range has been taken which is "50" in decimal then we converted it to hexadecimal which is "32". Therefore, the byte will be 0x32.

I. Moving the robot backward with fast speed

The command is shown in Fig. 16-c, used for moving the robot backward with fast speed. For this command, the fifth byte has been adjusted to be 0x9C. A negative value is chosen to move the robot backward. Therefore, the fifth byte represents the least two significant bits for -100 in decimal. The maximum speed has been used in a negative hexadecimal representation.

II. Moving the robot backward with medium speed

For moving the robot backward with medium speed, the command is shown in Fig. 16-d. The Fifth-byte value is decreased to be 0xCE in hexadecimal, which represents the least two bits -50 in decimal.

III. Move the robot to the left

This command is used to move the robot to the left. Fig. 16-e shows the structure of moving the robot to the left command. Moving all the motors doesn't require moving the robot to the left. For moving the robot to the left, motor B should move to the left and motor C should stop. Therefore, the value of the fourth byte will be modified to be 0x01 to select the output of motor B. Fifth byte will be 0x32 to move the robot with a medium speed.

IV. Moving the robot to the right

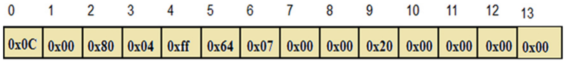
The command is shown in Fig. 16-f, used for moving the robot to the right, motor C should move to the right and motor B should stop. This command is similar to the command of moving the robot to the left, the difference is changing the fourth byte to be 0x02. So, motor C will be selected to be turned to the right.

V. Stop moving the robot

The last command is shown in Fig. 16-g is used to stop the robot's movement. All motors will be in an idle mood in this command. Thus, the fifth byte will be set to be 0x00.

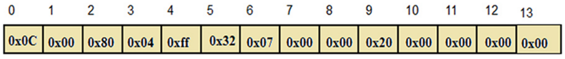
Table 1. The important commands with their types

The command	The type
0x00	Direct command telegram, response required
0x01	System command telegram, response required
0x02	Reply telegram
0x80	Direct command telegram, no response required



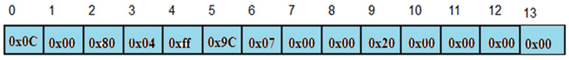
0	1	2	3	4	5	6	7	8	9	10	11	12	13
0x0C	0x00	0x80	0x04	0xff	0x64	0x07	0x00	0x00	0x20	0x00	0x00	0x00	0x00

(a) Forward Command telegram with Maximum Speed



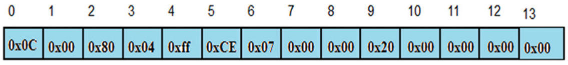
0	1	2	3	4	5	6	7	8	9	10	11	12	13
0x0C	0x00	0x80	0x04	0xff	0x32	0x07	0x00	0x00	0x20	0x00	0x00	0x00	0x00

(b) Forward Command Telegram with Medium Speed



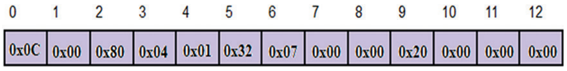
0	1	2	3	4	5	6	7	8	9	10	11	12	13
0x0C	0x00	0x80	0x04	0xff	0x9C	0x07	0x00	0x00	0x20	0x00	0x00	0x00	0x00

(c) Backward Command Telegram with Maximum Speed



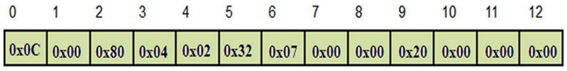
0	1	2	3	4	5	6	7	8	9	10	11	12	13
0x0C	0x00	0x80	0x04	0xff	0xCE	0x07	0x00	0x00	0x20	0x00	0x00	0x00	0x00

(d) Backward Command Telegram with Medium Speed



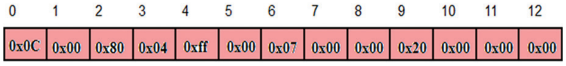
0	1	2	3	4	5	6	7	8	9	10	11	12
0x0C	0x00	0x80	0x04	0x01	0x32	0x07	0x00	0x00	0x20	0x00	0x00	0x00

(e) Moving Left Command Telegram



0	1	2	3	4	5	6	7	8	9	10	11	12
0x0C	0x00	0x80	0x04	0x02	0x32	0x07	0x00	0x00	0x20	0x00	0x00	0x00

(f) Moving Right Command Telegram



0	1	2	3	4	5	6	7	8	9	10	11	12
0x0C	0x00	0x80	0x04	0xff	0x00	0x07	0x00	0x00	0x20	0x00	0x00	0x00

(g) Stop Moving Command Telegram

Fig. 16. Different Command Telegram used for IoTR movement

3.3 REAL-TIME TRACKING SYSTEM

The tracing system is a fundamental sub-system in this robotic system as it is used for monitoring the movements of the patrol robot and live navigation for outdoor applications. Besides, the tracking system provides some other data such as current time, date, and the coordinates of the location during the movements of the robot. GPS module has been added to the hardware components for achieving a live tracking system. The GPS data will be sent to the web-server database through the GPRS connection. Then the user will be able to display the live location on Google Maps alongside the timing log data. The flowchart of the tracking system is shown in Fig. 8.

The GPS module that has been used in this system is EM-406. The GPS module is connected to the SparkFun GPS shield as shown in Fig. 17. The GPS Shield is connected at the top of the hardware implementation as shown in Fig. 18. The GPS shield has been connected with the GPRS shield using the pin assignment as given in Table 2. Pin 2 and 3 of the Arduino UNO are used only for the serial mode of the GPRS. To tackle this challenge, three steps have been accomplished. Firstly, the GPS shield should be connected at the top of the hardware system as these two pins can not be used for the serial mode of the GPS. Secondly, these two pins have been flexed to be out of connection with the Arduino. Thirdly, two other pins (9,10) will be plugged using two jumper wire connectors for the serial mode of the GPS, instead of the tucked GPS pins (1,3). Pin (2) will be connected with pin (9) of the GPS shield and pin (3) will be connected with pin (10).

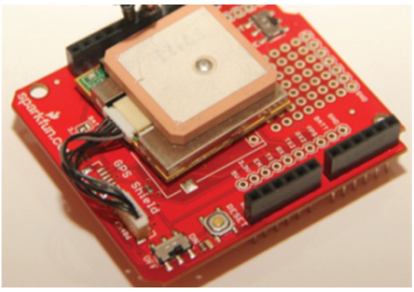


Fig. 17. GPS Module connected to the GPS Shield

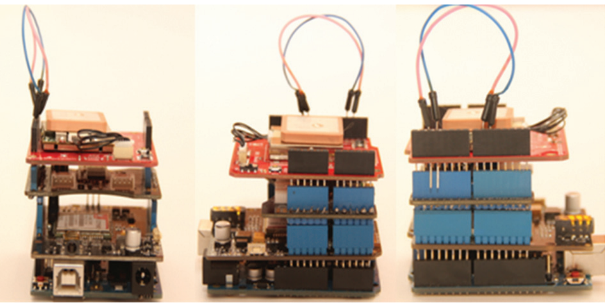


Fig. 18. GPS shield connected on the top of the hardware implementation

Table 2. EM- 406 GPS module pin assignment

Pin no.	Pin name	Usage
1	Enable/disable	For ON and OFF
2	GND	Provide the ground for the board
3	GPS-Rx	Receiving software commands
4	GPS-Tx	Outputting the measurement and navigation data to the user software
5	VIN	For DC supply

3.4 GAS LEAKAGE INSPECTION SYSTEM

The gas explosion occurs from a gas leakage with the excitement of an inflammation source [28]. The gas explosion happens with three conditions which are an explosive gas, an ignition source, and an oxidizer such as air or oxygen [29]. The most common explosive gases that are used for cooking and heating purposes are methane, butane, propane, and natural gas. In this proposed robotic system, LPG (Liquefied petroleum gas) has been the main explosive gas for testing and validation purposes. The reason for choosing LPG is this gas is available in the market and can be tested for personal and educational usage. While other gases such as natural gas and propane need a special safe environment for testing. LPG is propane or butane or a mixture of 60% propane and 40% butane that is used for commercial use. LPG gas is highly used for heating in homes or hotels or vehicles, however, LPG gas leakage is explosive and flammable [30].

For the LPG, the flame point is considered almost at the same point as the butane at 1970 °C in the air condition. In this research, we are not able to measure the flame point or the level of danger in terms of temperature, but we used a specific sensor to detect the LPG gas under certain conditions which is the MQ-6 gas sensor. There are different types of gas sensors in the market. These gas sensors have a small heater with an electrochemical sensor [31]. Besides, they are used at room temperature and outdoors, below are four different gas and chemical sensors that have been tested and validated.

A. MQ-5 Gas Sensor

MQ-5 sensor can be used in several types of applications such as "Domestic gas leakage detector, industrial combustible gas detector, and portable gas detector" [32]. This gas sensor is used to detect the leakages of natural gases.

B. MQ-6 Gas Sensor

MQ6 has been used due to two reasons which are low conductivity in clean air and low sensitivity to cigarette smoke and alcohol. Therefore, it is easy for testing and validation. Moreover, the MQ-6 gas sensor has a fast response time which is less than 10s [33].

C. ChemSee's Nitro-Pen Sensor

Trinitrotoluene (TNT) is a popular explosive material that has been used in many military and industry applications [34]. ChemSee's Nitro-Pen Sensor is a very effective chemical detector that can be built into the patrol robot to detect TNT. Fig. 19 shows the TNT testing using that detector.

D. Nanotechnology Chemical Sensor

A new chemical sensor chip joint with carbon Nanotubes has been made[35]. It enables the rapid detection of TNT in rivers and reservoirs. Fig. 20 shows the chemical sensor chip using Nanotechnology and the detection of the TNT process. This sensor is used by semiconducting carbon nanotube network transistors to make extremely sensitive sensors that are capable of operating stably underwater. Therefore, it is sensitive in the range of a few parts per billion for the detection of explosive compounds such as TNT [35]. Fig. 21 shows a diagram of a Nano-tube transistor on a flexible chip for detecting toxins or explosives in a water sample.

Based on experiments and testing, this research focused on using two sensors which are MQ5 and MQ6. The usage of MQ5 and MQ6 has the advantages of its low cost, availability in the market, and is lightweight compared to other gas sensors. On the other hand, these types of sensors have a lack of selectivity and cross-sensitivity to environmental factors such as temperature and humidity. In order to tackle this challenge, two sensors have been used, one for operating different temperature measurements and the other one for humidity measurements. These gas sensors obtain fast and calibrated measurements for concertation type of gases. The main hardware components that are used for hazard gases detection are explained in the following subsections.

3.4.1 MQ6 GAS SENSOR

MQ-6 gas sensor has a high sensitivity to LPG, Iso-butane, and propane gas. MQ6 gas sensor has a shield that is compatible with the Arduino shield. This shield of MQ6 has 3 pins which are Vcc, GND, and Vout. These 3 pins have connected to the Arduino shield; The Vcc pin of the MQ6 sensor is connected to +5v of the Arduino shield. The GND pin of the MQ6 sensor is connected to the GND of the Arduino shield. The Vout pin of the MQ6 sensor is connected to the A0 pin, which represents analog input to the Arduino. Fig. 22 shows the schematic circuit for the MQ6 sensor connected to the Arduino shield.

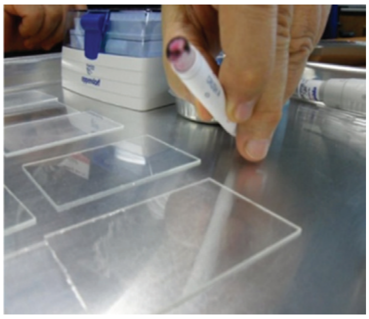


Fig. 19. TNT Testing using ChemSee's Nitro-pen

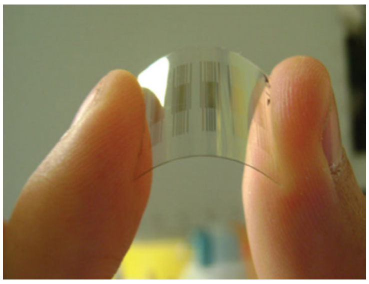


Fig. 20. The Chemical Sensor Chip using Nanotechnology and detection of TNT process

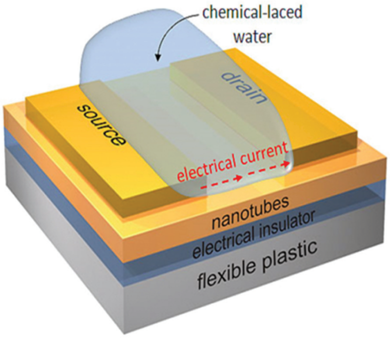


Fig. 21. The diagram of a Nano-tube transistor on a flexible chip for detecting toxins or explosives in the water

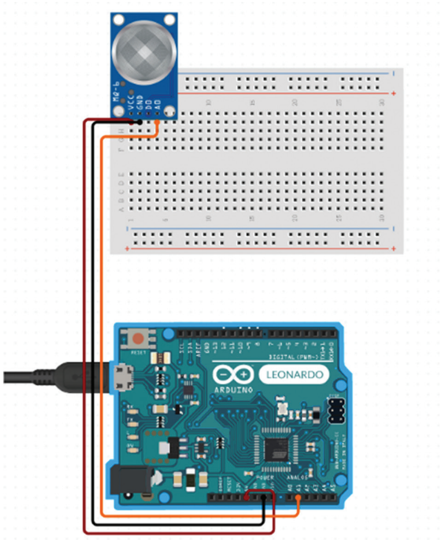


Fig. 22. Schematic circuit for MQ6 Gas sensor connected to Arduino UNO

The sensor is connected at the top of the GPS shield as shown in Fig. 23. According to [36], the sensor provides a reading of 100mv in the clean air, then the output voltage starts to increase according to the increase of the gas concertation. The resistance ratio of the sensor is calculated using Rs/Ro. Rs means the resistance of the sensor in 1000ppm methane under different temperatures and humidity. Ro means the resistance of the sensor in the environment of 1000ppm propone. Rs calculated as in Equation (1).

$$Rs = (\frac{v_c}{v_{RL} - 1}) * RL \qquad (1)$$

V_{RL} is the voltage across the load resistance which is the voltage output of the MQ6. V_c is the reference voltage which is equal to 5v. RL is the load resistance which is 10 kΩ. PPM value of the LPG concertation is calculated using Equation (2).

$$PPM = \sqrt[-0.421]{\frac{Rs}{Ro * 18.446}} \qquad (2)$$

The calibration of the sensor has been done with the collaboration of the HPC company team. Based on experiments and testing, the threshold level of the gas sensor result is 750mv to be considered a dangerous level.

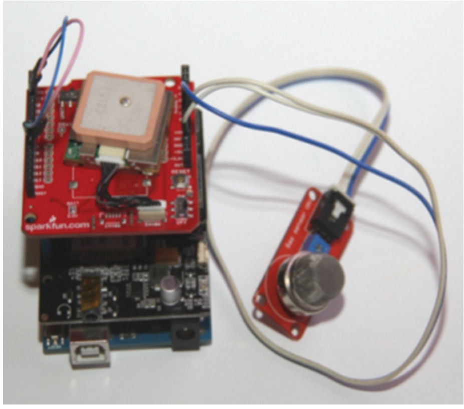


Fig. 23. Gas detection using MQ-6 Gas Sensor and Interfacing to System Board

3.4.2 TEMPERATURE SENSOR

The proposed gas sensor that has been used, worked within a certain condition of temperature which is in a range of -10C to 70C. This was the reason to use a temperature sensor to measure the temperature in the tested environment if it meets this condition. The temperature sensor used for that purpose is LM35. It works in a temperature range between -55C to 150C [37]. This range will contain the range of the gas sensor condition. LM35 has three pins to be connected. The first pin from the left is connected to the VCC of the Arduino. The second pin is the Vout. Vout is connected to resistance that equals 4.7 KΩ and it is connected to the analog input pin A2 of the Arduino. The third pin is connected to the common GND in the Arduino chip. Fig. 24 shows the schematic design of the LM35 sensor with the Arduino shield.

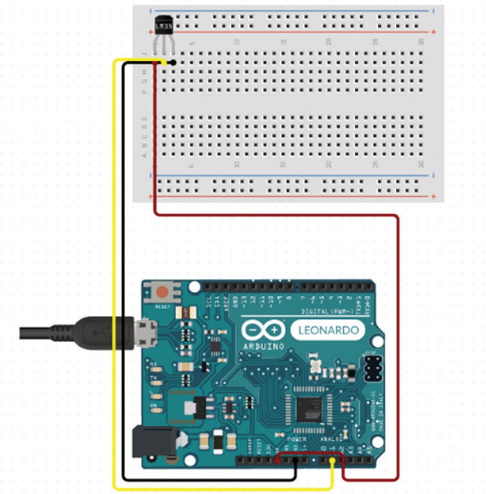


Fig. 24. Schematic design of the LM35 connected to Arduino shield

Connecting this sensor in the proposed system will be on the top of the GPS shield as shown in Fig. 25. The 4.7 KΩ resistance is used for the safety of the circuit and to avoid burning of the temperature sensor. The output value obtained from the analog pin will be converted to mv and then it will be analyzed and converted to Celsius degrees in the software development code using C language.

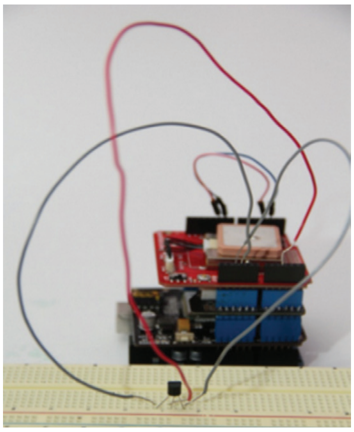
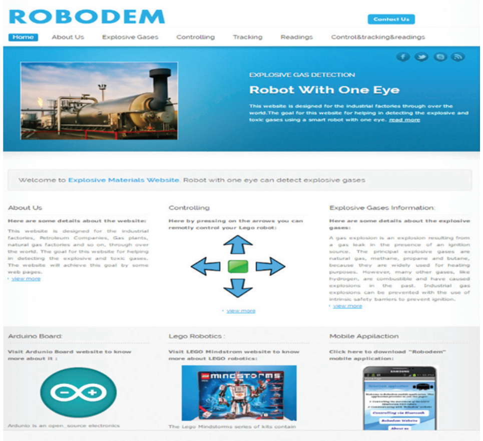


Fig. 25. LM35 Temperature Sensor connected to the system

4. SOFTWARE USER INTERFACE

Patrol robot communicates with mobile apps or web-server through GPRS network to send or receive commands. The default setting of the baud rate for the GPRS is 19200. Table 3 displays the AT commands that have been used for the connection of GPRS [9]. The web-based system software with smart applications has been designed and tested, and it is available online, a screenshot of the website system is shown in Fig. 26. Additionally, the mobile application has been developed and tested to achieve the main four sub-systems of the patrol robot. Some of the main pages of the mobile application are shown in Fig. 27. The users have the flexibility to use both of them based on their preferences. Fig. 28 shows the flow chart for the mobile application of IoTR.



EXPLOSIVE GAS DETECTION

Robot With One Eye

This website is designed for the industrial factories through over the world.The goal for this website for helping in detecting the explosive and toxic gases using a smart robot with one eye. read more

Welcome to Explosive Materials Website. Robot with one eye can detect explosive gases

About Us

Here are some details about the website:

This website is designed for the industrial factories, Petroleum Companies, Gas plants, natural gas factories and so on, through over the world. The goal for this website for helping in detecting the explosive and toxic gases. The website will achieve this goal by some web pages.

- view more

Controlling

Here by pressing on the arrows you can remotly control your Lego robot:

- view more

Explosive Gases Information:

Here are some details about the explosive gases:

A gas explosion is an explosion resulting from a gas leak in the presence of an ignition source. The principal explosive gases are natural gas, methane, propane and butane, because they are widely used for heating purposes. However, many other gases, like hydrogen, are combustible and have caused explosions in the past. Industrial gas explosions can be prevented with the use of intrinsic safety barriers to prevent ignition.

- view more

Arduino Board:

Visit Ardunio Board website to know more about it :

Ardunio is an open_source electronics

Lego Robotics :

Visit LEGO Mindstrom website to know more about LEGO robotics:

The Lego Mindstorms series of kits contain

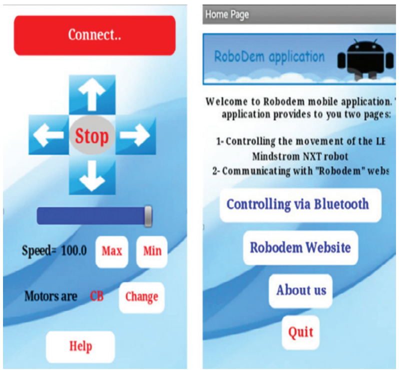
Mobile Appilaction

Click here to download "Robodem" mobile application:

Fig. 26. The interface for the IoTR website

Table 3. AT commands used for the GPRS

The AT command	The meaning
AT+CSQ	Check Signal quality
AT+CGATT	Check the status of Packet service attach
AT+SAPBR	Bearer settings for applications based on IP
AT+HTTPINIT	Initiate the HTTP request



Welcome to Robodem mobile application. application provides to you two pages:

1- Controlling the movement of the LE Mindstrom NXT robot

2- Communicating with "Robodem" webs

Fig. 27. Some pages of IoTR Mobile Application

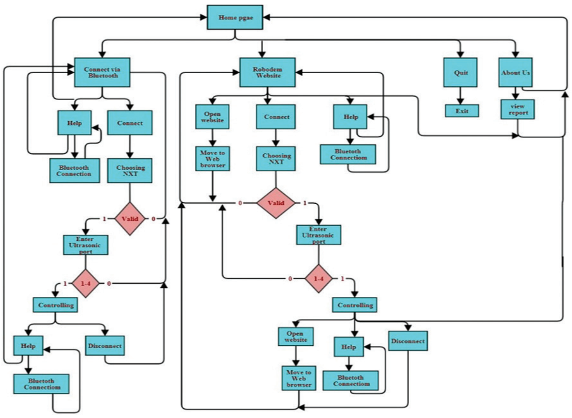


Fig. 28. Flow chart for mobile application of IoTR

5. TESTING AND VALIDATION RESULTS

This section presents the experimental results in terms of testing and validation for the whole system and subsystems. Different experiments have been conducted mostly at the University of Bahrain campus whether indoor or outdoor. Testing web-server and mobile apps verified successfully by sending and receiving the data from and into the database. The full hardware implementation of the IoTR prototype is shown in Fig. 29, which contains all the hardware units and sensors. The testing and validation results of each subsystem have been discussed in the following subsections.

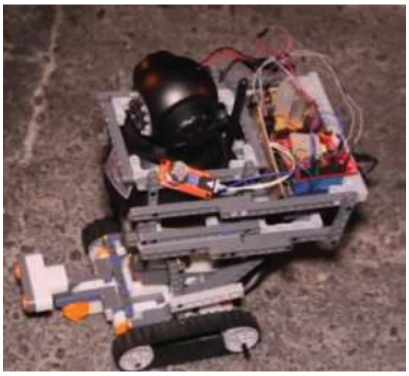


Fig. 29. IoT Robot

5.1 OUTDOOR CONTROLLING RESULTS

The IoT robot has been tested for outdoor controlling subsystem as shown in Fig. 30. The purpose of testing the outdoor controlling is to verify the output of the software development code and the hardware implementation by getting a fast response from IoTR. For the hardware implementation of the controlling system, the hardware units which are the Arduino UNO board, GPRS shield, and the Bluetooth shield have been connected. Then, the software code for controlling is uploaded and run. The results have been shown on the serial mode that obtains the data log of the controlling. Verifying the setup of the GPRS connection is an important step, as the GPRS is responsible for receiving the data from the database of the web-server and sending it back to the mobile robot to make a successful controlling through Bluetooth. After making a connection between the Arduino shield and the GPRS shield, the serial mode has been shown the connection data of the GPRS as in Fig. 31. The serial mode showed the SAPBR connection type of the GPRS is established. Also, it verified the local APN server which is the 'VIVA' network provider that has been used. After that, the AT command and the HTTP request have been successfully sent.

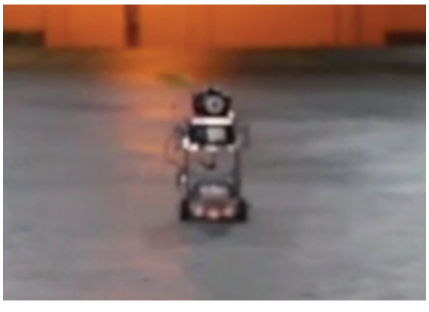
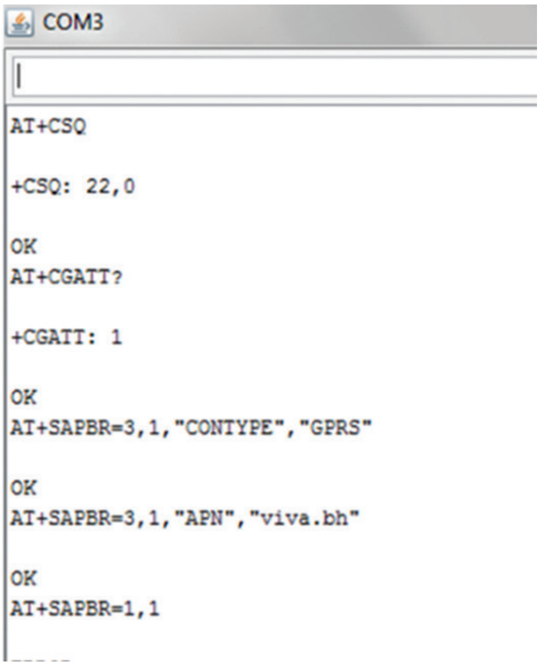


Fig. 30. Testing the Outdoor Controlling System



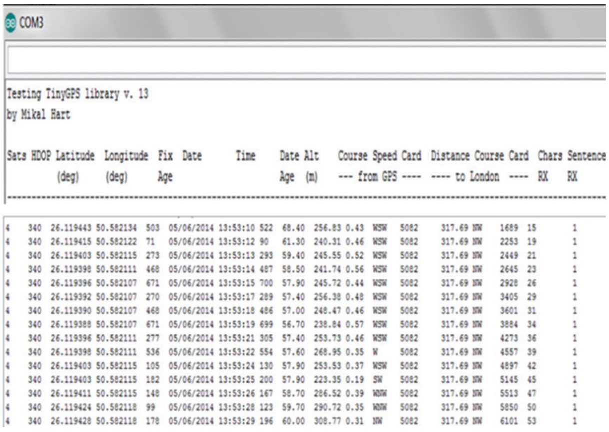
```
AT+CSQ

+CSQ: 22,0

OK
AT+CGATT?

+CGATT: 1

OK
AT+SAPBR=3,1,"CONTYPE","GPRS"

OK
AT+SAPBR=3,1,"APN","viva.bh"

OK
AT+SAPBR=1,1
```

Fig. 31. Setting up the GPRS for outdoor controlling

The GPRS can transmit data outdoor at unlimited distances. Therefore, the users can remotely control the robot from anywhere to avoid accidents and save their lives in harsh environments. The data transmission speed (baud rate) used is 19200 bits per second (bps).

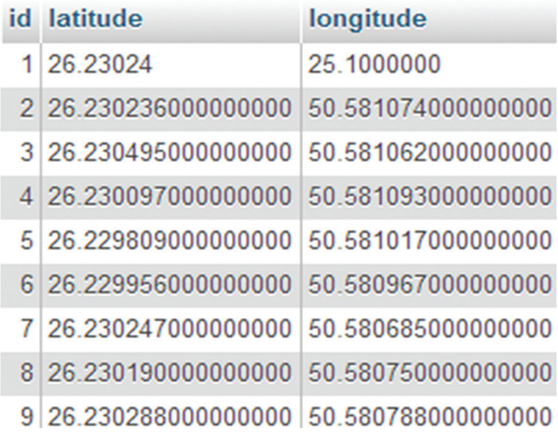
5.2 TRACKING SYSTEM RESULTS

The main purpose of testing the GPS shield is to verify the tracking and the localization of the IoT robot. Firstly, the GPS shield has been connected to the hardware units. After a few seconds, accurate GPS data has been received which are the date, time, longitude, and latitude of the location. Fig. 32 shows the GPS data and results on the serial mode. The tracking system has been tested to trigger the last location reporting and real-time tracking. Fig. 33 shows the results of the last location and reading history through the website. The live tracking has been tested at different locations when the robot moves. Fig. 34 and Fig. 35 show two different live locations for the movement of the robot.



```
Testing TinyGPS library v. 13
by Mikal Hart

Sats HDOP Latitude  Longitude  Fix  Date       Time     Date Alt    Course Speed Card  Distance Course Card  Chars Sentence
          (deg)     (deg)      Age                      Age  (m)    --- from GPS ----  ---- to London  ----  RX    RX
------------------------------------------------------------------------------------------------------------------------------

4    340  26.119443 50.582134  503  05/06/2014 13:53:10 522  68.40  256.83 0.43  WSW   5082     317.69 NW    1689  15     1
4    340  26.119415 50.582122  71   05/06/2014 13:53:12 90   61.30  240.31 0.46  WSW   5082     317.69 NW    2253  19     1
4    340  26.119403 50.582115  273  05/06/2014 13:53:13 293  59.40  245.55 0.52  WSW   5082     317.69 NW    2449  21     1
4    340  26.119398 50.582111  468  05/06/2014 13:53:14 487  58.50  241.74 0.56  WSW   5082     317.69 NW    2645  23     1
4    340  26.119396 50.582107  671  05/06/2014 13:53:15 700  57.90  245.72 0.44  WSW   5082     317.69 NW    2928  26     1
4    340  26.119392 50.582107  270  05/06/2014 13:53:17 289  57.40  256.38 0.48  WSW   5082     317.69 NW    3405  29     1
4    340  26.119390 50.582107  468  05/06/2014 13:53:18 486  57.00  248.47 0.46  WSW   5082     317.69 NW    3601  31     1
4    340  26.119388 50.582107  671  05/06/2014 13:53:19 699  56.70  238.84 0.57  WSW   5082     317.69 NW    3884  34     1
4    340  26.119396 50.582111  277  05/06/2014 13:53:21 305  57.40  253.73 0.46  WSW   5082     317.69 NW    4273  36     1
4    340  26.119398 50.582111  536  05/06/2014 13:53:22 554  57.60  268.95 0.35  W     5082     317.69 NW    4557  39     1
4    340  26.119403 50.582115  105  05/06/2014 13:53:24 130  57.90  253.53 0.37  WSW   5082     317.69 NW    4897  42     1
4    340  26.119403 50.582115  182  05/06/2014 13:53:25 200  57.90  223.35 0.19  SW    5082     317.69 NW    5145  45     1
4    340  26.119411 50.582115  148  05/06/2014 13:53:26 167  58.70  286.52 0.39  WNW   5082     317.69 NW    5513  47     1
4    340  26.119424 50.582118  99   05/06/2014 13:53:28 123  59.70  290.72 0.35  WNW   5082     317.69 NW    5850  50     1
4    340  26.119428 50.582118  178  05/06/2014 13:53:29 196  60.00  308.77 0.31  NW    5082     317.69 NW    6101  53     1
```

Fig. 32. The GPS data results on serial mode



id	latitude	longitude
1	26.23024	25.1000000
2	26.230236000000000	50.581074000000000
3	26.230495000000000	50.581062000000000
4	26.230097000000000	50.581093000000000
5	26.229809000000000	50.581017000000000
6	26.229956000000000	50.580967000000000
7	26.230247000000000	50.580685000000000
8	26.230190000000000	50.580750000000000
9	26.230288000000000	50.580788000000000

Fig. 33. The last location of the robot on Google map and the reading results of the GPS

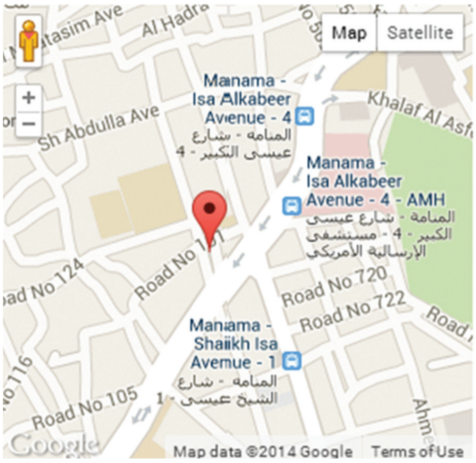

Map
Satellite
Manama - Isa Alkabeer Avenue - 4
المنامة - شارع عيسى الكبير - 4
Manama - Isa Alkabeer Avenue - 4 - AMH
المنامة - شارع عيسى الكبير - 4 - مستشفى الإرسالية الأمريكية
Manama - Shaikh Isa Avenue - 1
المنامة - شارع الشيخ عيسى - 1
Sh Abdulla Ave
Khalaf Al Asf
Road No 124
Road No 1
Road No 720
Road No 722
Road No 105
Al Hadra
Google
Map data ©2014 Google
Terms of Use

Fig. 34. Live Tracking - Test location -1

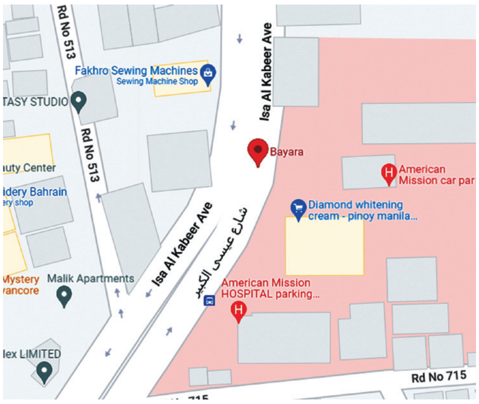
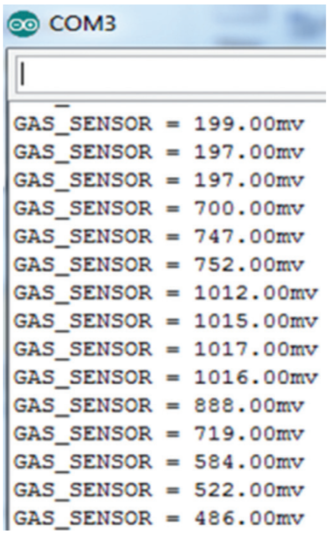


Fig. 35. Live Tracking - Test location -2

5.3 GAS SENSORS RESULTS

Firstly, the gas sensor (MQ6) is tested using a voltammeter. The results obtained from the voltammeter vary from 100 to 1200 mv. Secondly, the MQ-6 sensor has a high sensitivity to Propane, Butane, and LPG. Therefore, the testing has been done using a cigarette lighter or smoker because it contains a percentage of butane. The temperature sensor has been tested by using a candle. After that, the voltage readings are increased when the smoker cigarette gets closer to the sensor. On the other hand, the voltage readings decreased slightly when the cigarette lighter got away from the sensor. All of these testing scenarios mean that the changing of the voltage readings depended on the concentration of the ppm of the gases. The concentration of the gases should be between 200-10000ppm. Fig. 36 shows the results of the MQ6 sensor. The change of LPG concertation and the temperature values have been validated under the same measurements. Fig. 37 shows the LPG monitoring chart. Fig. 37 Illustrates that the threshold level is 750 mV. If the LPG ppm concertation is less than the threshold which is 750 mv, it means the surrounding area is safe and contains fresh air. On the other hand, if the LPG ppm concertation is equal to or more than 750 mv, it means there is a leakage of hazardous gases that will cause accidents and explosions. The reading of smart sensors has been verified through the database storage. Besides, the system generates user-friendly dashboard visualization reports for the gases and temperature sensors readings based on filtration by the date and time. Fig. 38 shows the sensor's results with showing the red highlighted row that represents a gas leakage. Finally, the system sends a warning message and a sign of gas leaks through the website and mobile application once there is a trigger of a gas leakage as shown in Fig. 39. Moreover, if the concentration of the explosive gas is getting higher, the Piezo buzzer will make an alarming sound.



```
GAS_SENSOR = 199.00mv
GAS_SENSOR = 197.00mv
GAS_SENSOR = 197.00mv
GAS_SENSOR = 700.00mv
GAS_SENSOR = 747.00mv
GAS_SENSOR = 752.00mv
GAS_SENSOR = 1012.00mv
GAS_SENSOR = 1015.00mv
GAS_SENSOR = 1017.00mv
GAS_SENSOR = 1016.00mv
GAS_SENSOR = 888.00mv
GAS_SENSOR = 719.00mv
GAS_SENSOR = 584.00mv
GAS_SENSOR = 522.00mv
GAS_SENSOR = 486.00mv
```

Fig. 36. Gas Sensor Results

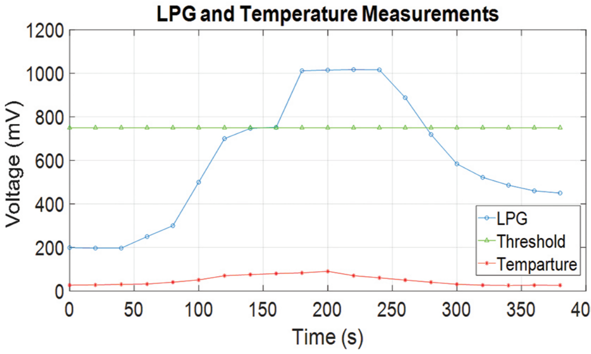
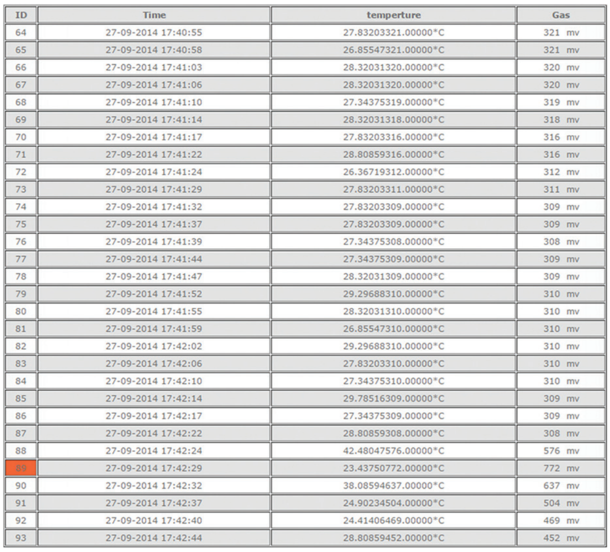


Fig. 37. LPG and Temperature Measurements



ID	Time	temperture	Gas
64	27-09-2014 17:40:55	27.83203321.00000*C	321 mv
65	27-09-2014 17:40:58	26.85547321.00000*C	321 mv
66	27-09-2014 17:41:03	28.32031320.00000*C	320 mv
67	27-09-2014 17:41:06	28.32031320.00000*C	320 mv
68	27-09-2014 17:41:10	27.34375319.00000*C	319 mv
69	27-09-2014 17:41:14	28.32031318.00000*C	318 mv
70	27-09-2014 17:41:17	27.83203316.00000*C	316 mv
71	27-09-2014 17:41:22	28.80859316.00000*C	316 mv
72	27-09-2014 17:41:24	26.36719312.00000*C	312 mv
73	27-09-2014 17:41:29	27.83203311.00000*C	311 mv
74	27-09-2014 17:41:32	27.83203309.00000*C	309 mv
75	27-09-2014 17:41:37	27.83203309.00000*C	309 mv
76	27-09-2014 17:41:39	27.34375308.00000*C	308 mv
77	27-09-2014 17:41:44	27.34375309.00000*C	309 mv
78	27-09-2014 17:41:47	28.32031309.00000*C	309 mv
79	27-09-2014 17:41:52	29.29688310.00000*C	310 mv
80	27-09-2014 17:41:55	28.32031310.00000*C	310 mv
81	27-09-2014 17:41:59	26.85547310.00000*C	310 mv
82	27-09-2014 17:42:02	29.29688310.00000*C	310 mv
83	27-09-2014 17:42:06	27.83203310.00000*C	310 mv
84	27-09-2014 17:42:10	27.34375310.00000*C	310 mv
85	27-09-2014 17:42:14	29.78516309.00000*C	309 mv
86	27-09-2014 17:42:17	27.34375309.00000*C	309 mv
87	27-09-2014 17:42:22	28.80859308.00000*C	308 mv
88	27-09-2014 17:42:24	42.48047576.00000*C	576 mv
89	27-09-2014 17:42:29	23.43750772.00000*C	772 mv
90	27-09-2014 17:42:32	38.08594637.00000*C	637 mv
91	27-09-2014 17:42:37	24.90234504.00000*C	504 mv
92	27-09-2014 17:42:40	24.41406469.00000*C	469 mv
93	27-09-2014 17:42:44	28.80859452.00000*C	452 mv

Fig. 38. The dashboard report of the gas sensor and temperature results

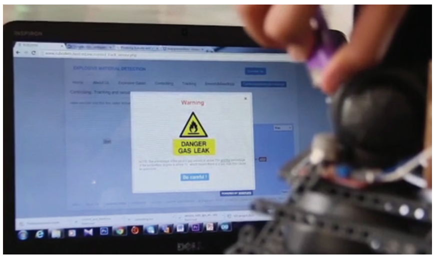


Fig. 39. A warning message appeared at the user-interface

6. CONCLUSIONS AND FUTURE WORK

This work describes the development and hardware implementation of a low-cost real-time IoT robot that can be used either indoors or outdoors for surveillance and gas leakage inspection purposes. The system has many features that are used by the security and gas detection system to locate the robot using GPS/GPRS for outdoor applications or Bluetooth for indoor applications. The design approach is economically better and has real-time features when compared to other approaches in the state-of-the-art related work. The proposed design achieves a low-cost and high quality, it has been tested indoors and outdoors for real-time monitoring, gas leakage inspection, remote controlling, live tracking, and video streaming. The developed autonomous mobile robot has the obstacle avoidance feature, which is a common feature of robotic patrolling. However, an optimized algorithm for obstacle detection is planned for future work. The hardware and software design of the gas detection system has been validated and tested using intelligent gas sensors. LPG gas has been sensed using MQ6 gas sensor. More different gases will be tested in future work. The network protocols between the microcontroller of the robot and the web-server have been established and tested for three main features which are controlling using the website and the mobile application, live video streaming using IP cameras, and live tracking on google maps.

7. REFERENCES

[1] D.-M. Tsai, T.-H. Tseng, "A template reconstruction scheme for moving object detection from a mobile robot", Industrial Robot: An International Journal, 2013, pp. 559-573.

[2] D. Xu, X. Peng, "Multi-level offloading decision on efficient object tracking for humanoid robot", Advanced Robotics, Vol. 30, No. 15, 2016, pp. 979-991.

[3] A. Shukla, H. Karki, "Application of robotics in onshore oil and gas industry—A review Part I", Robotics and Autonomous Systems, Vol. 75, 2016, pp. 490-507.

[4] D. A. Anisi, J. Gunnar, T. Lillehagen, C. Skourup, "Robot automation in oil and gas facilities: Indoor and onsite demonstrations", Proceedings of the IEEE RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18-22 October 2010, pp. 4729-4734.

[5] M. Choyekh et al. "Development and Operation of Underwater Robot for Autonomous Tracking and Monitoring of Subsea Plumes After Oil Spill and Gas Leak from Seabed and Analyses of Measured Data", Applications to Marine Disaster Prevention, 2017, pp. 17-93.

[6] Y. Li, Q. Liu, W. Li, "Development of a novel oil and gas in-pipe robot", International Journal of Mechatronics and Manufacturing Systems, Vol. 8, No. 3-4, 2015, pp. 102-115.

[7] Y. Tipsuwan, P. Hoonsuwan, "Design and implementation of an AUV for petroleum pipeline inspection", Proceedings of the 7th International Conference on Information Technology and Electrical Engineering, Chiang Mai, Thailand, 29-30 October 2015, pp. 382-387.

[8] T. M. Yousif, A. K. Alharam, W. Elmedany, A. A. Al Khalaf, Z. Fardan, "ROBODEM Remote Monitoring System Using Web/Mobile Applications", Proceedings of the Fifth International Conference on e-Learning, Manama, Bahrain, 18-20 October 2015, pp. 285-289.

[9] T. M. Yousif, A. K. Alharam, W. Elmedany, A. A. AlKhalaf, Z. Fardan, "GPRS-based robotic tracking system with real time video streaming", Proceedings of the IEEE 4th International Conference on Future Internet of Things and Cloud Workshops, Vienna, Austria, 22-24 August 2016, pp. 299-303.

[10] M. Sharma, H. Elmiligi, F. Gebali, "Performance evaluation of real-time systems", International Journal of Computing and Digital Systems, Vol. 4, No. 1, 2015.

[11] D. M. Alghasra, H. Y. Saeed, "Guiding Visually Impaired People with NXT Robot through an Android Mobile Application", International Journal of Computing and Digital Systems, Vol. 2, No. 3, 2013.

[12] S. Soldan, J. Welle, T. Barz, A. Kroll, D. Schulz, "Towards autonomous robotic systems for remote gas leak detection

and localization in industrial environments", Springer Field and service robotics, 2014, pp. 233-247.

[13] M. Faisal, H. Mathkour, M. Alsulaiman, "Smart mobile robot for security of low visibility environment", Proceedings of the 5th National Symposium on Information Technology: Towards New Smart World, Riyadh, Saudi Arabia, 17-19 February 2015, pp. 1-6.

[14] J. Patoliya, H. Mehta, H. Patel, "Arduino controlled war field spy robot using night vision wireless camera and Android application", Proceedings of the 5th Nirma University International Conference on Engineering, Ahmedabad, India, 26-28 November 2015, pp. 1-5.

[15] V. A. Shadrin, S. A. Lisakov, A. N. Pavlov, E. V. Sypin, "Development of fire robot based on quadcopter", Proceedings of the 17th International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, Erlagol, Russia, 30 June - 4 July 2016, pp. 364-369.

[16] K. S. Lee, M. Ovinis, T. Nagarajan, R. Seulin, O. Morel, "Autonomous patrol and surveillance system using unmanned aerial vehicles", Proceedings of the IEEE 15th International Conference on Environment and Electrical Engineering, Rome, Italy, 10-13 June 2015, pp. 1291-1297.

[17] W. Rahmaniar, A. Wicaksono, "Design and implementation of a mobile robot for carbon monoxide monitoring", Journal of Robotics and Control, Vol. 2, No. 1, 2021, pp. 1-6.

[18] J. Palacín et al., "Application of an array of metal-oxide semiconductor gas sensors in an assistant personal robot for early gas leak detection", Sensors, Vol. 19, No. 9, 2019, p. 1957.

[19] T. Ma, S. Liu, H. Xiao, "Location of natural gas leakage sources on offshore platform by a multi-robot system using particle swarm optimization algorithm", Journal of Natural Gas Science and Engineering, Vol. 84, 2020, p. 103636.

[20] Y. Xing et al., "FireNose on mobile robot in harsh environments", IEEE Sensors Journal, Vol. 19, No. 24, 2019, pp. 12418-12431.

[21] S. Soldan, J. Welle, T. Barz, A. Kroll, D. Schulz, "Towards autonomous robotic systems for remote gas leak detection and localization in industrial environments", Springer Field and Service Robotics, 2014, pp. 233-247.

[22] D. S. Moore, R. J. Scharff, "Portable Raman explosives detection", Analytical and bioanalytical chemistry, Vol. 393, No. 6, 2009, pp. 1571-1578.

[23] M. Ghazal, R. Hamouda, S. Ali, "A smart mobile system for the real-time tracking and management of service queues", International Journal of Computing and Digital Systems, Vol. 5, No. 04, 2016.

[24] Y. Ismail, "A fast diamond motion estimation search algorithm for real time video applications", International Journal of Computing and Digital Systems, Vol. 3, No. 2, 2014, pp. 101-110.

[25] C. Bouras, A. Papazois, N. Stasinos, "Cross-platform Mobile Applications with Web Technologies", International Journal of Computing and Digital Systems, Vol. 4, No. 03, 2015.

[26] M. Eltoweissy, W. El-Medany, "Introduction to Special Issue on Mobile Applications", International Journal of Computing and Digital Systems, Vol. 5, No. 01, 2016.

[27] LEGO MINDSTORM, LEGO MINDSTORM NXT Direct Commands, http://remy-manu.no-ip.biz/Java/Tutoriels/Robot/ressources/LEGO_MINDSTORMS_NXT_Direct_commands.pdf (accessed: 2022)

[28] P. X. Thivel, Y. Bultel, F. Delpech, "Risk analysis of a biomass combustion process using MOSAR and FMEA methods", Journal of hazardous materials, Vol. 151, No. 1, 2008, pp. 221-231.

[29] V. Schröder, K. Holtappels, "Explosion characteristics of hydrogen-air and hydrogen-oxygen mixtures at elevated pressures", Hydrogen Knowledge Centre, 2005.

[30] H. Fayaz, R. Saidur, N. Razali, F. S. Anuar, A. R. Saleman, M. R. Islam, "An overview of hydrogen as a vehicle fuel", Renewable and Sustainable Energy Reviews, Vol. 16, No. 8, 2012, pp. 5511-5528.

[31] H. Maurer, F. Rieger, E. Linder, "Electro-chemical sensor construction," Modern Environmental Analysis Techniques for Pollutants, USA, Technical Report TR-0572595, 1979.

[32] J. Sun, Y. Li, X. Yan, "The design of automatic detection processing device of gas leakage based on the MB95204K", IEEE International Conference on Electrical and Control Engineering, 2011, pp. 1807-1809.

[33] A. Shrivastava, R. Prabhaker, R. Kumar, R. Verma, "GSM based gas leakage detection system", International Journal of Technical Research and Applications, Vol. 1, No. 2, 2013, pp. 42-45.

[34] S. J. Toal, J. C. Sanchez, R. E. Dugan, W. C. Trogler, "Visual detection of trace nitroaromatic explosive residue using photoluminescent metallole-containing polymers", Journal of forensic sciences, Vol. 52, No. 1, 2007, pp. 79-83.

[35] P. Shabecoff, A. Shabecoff, "Poisoned profits: The toxic assault on our children", 2nd Edition, Random House, 2008.

[36] HANWEI Sensors, Technical Data MQ6 Gas Sensor, https://datasheetspdf.com/pdf-file/699271/HANWEI/MQ6/1 (accessed: 2021)

[37] J. Akhavan, "The chemistry of explosives", 4th Edition, Royal Society of Chemistry, 2011.

A Trust-Based Recommender System for Personalized Restaurants Recommendation

Original Scientific Paper

Qusai Shambour
Al-Ahliyya Amman University
Faculty of Information Technology
Amman, Jordan
q.shambour@ammanu.edu.jo

Mosleh M. Abualhaj
Al-Ahliyya Amman University
Faculty of Information Technology
Amman, Jordan
m.abualhaj@ammanu.edu.jo

Ahmad Adel Abu-Shareha
Al-Ahliyya Amman University
Faculty of Information Technology
Amman, Jordan
a.abushareha@ammanu.edu.jo

Abstract *– Several online restaurant applications, such as TripAdvisor and Yelp, provide potential consumers with reviews and ratings based on previous customers' experiences. These reviews and ratings are considered the most important factors that determine the customer's choice of restaurants. However, the selection of a restaurant among many unknown choices is still an arduous and time-consuming task, particularly for tourists and travellers. Recommender systems utilize the ratings provided by users to assist them in selecting the best option from many options based on their preferences. In this paper, we propose a trust-based recommendation model for helping consumers select suitable restaurants in accordance with their preferences. The proposed model utilizes multi-criteria ratings of restaurants and implicit trust relationships among consumers to produce personalized restaurant recommendations. The experimental results based on a real-world restaurant dataset demonstrated the superiority of the proposed model, in terms of prediction accuracy and coverage, in overcoming the sparsity and new user problems when compared to other baseline CF-based recommendation algorithms.*

Keywords*: restaurant, collaborative filtering, recommender systems, multi-criteria, sparsity, new user*

1. INTRODUCTION

The restaurant industry has experienced a remarkable expansion in recent years, with a slew of new restaurants springing up. Consumers nowadays are more interested in trying a variety of cuisines, and just because there are many restaurants does not mean that everyone will visit each one and try everything. Moreover, the internet offers a vast amount of online restaurant review information, which is particularly useful for consumers when deciding where to dine. Consumers, on the other hand, frequently find it time-consuming and difficult to extract meaningful information from the vast amount of online information available, making selecting a restaurant even more challenging. Given how people's lifestyles are changing as a result of their use of technology, an intelligent recommender system that recommends restaurants can be a good solution for consumers to assist them in finding a restaurant that fits their needs and preferences [1-7].

Recommender systems are programs that aim to suggest the most appropriate items (services and products) to specific users by anticipating a user's interest in an item based on relevant information about the items, the users, and their interactions with the items. The goal of building recommender systems is to avoid information overload by obtaining the most relevant information from a large amount of data, allowing personalized services to be provided in various disciplines such as e-commerce, e-business, e-library, e-learning, e-tourism, e-health, and e-government [8-13]. Collaborative filtering (CF) is a well-recognized technique in recommender systems for producing personalized recommendations based on: 1) rating information of items liked by other similar users (user-based CF), and 2) rating information of items similar to items the user has liked in the past (item-based CF). User-based CF approaches have proven to be effective recommendation approaches across a variety of disciplines. However, due to major constraints such as sparsity and new user, they may provide poor recommendations (reducing accuracy) or decline recommendations (reducing coverage). Sparsity arises from a lack of user ratings, particularly when the number of ratings obtained is modest in comparison to the number of ratings that must be predicted. The term "new user" refers to a user who has only rated a very small number of items [14]. To address these issues and enhance the accuracy and coverage of user-based CF recommender systems, researchers have lately begun to utilize the multi-criteria ratings of users and combine CF with additional information, such as the trust relationships between users, to provide more trustworthy recommendations [15,16].

To generate recommendations, most current recommender systems solely consider an item's overall rating, which is a single-criterion score from a user. However, it has been established that while selecting an item, a user may consider more than one feature of the item. That is, understanding why users like things in addition to what they like is critical to making more successful recommendations. To put it another way, the exploitation of multi-criteria ratings of items can ensure a better understanding of users' preferences, hence improving recommendation accuracy [15-17]. Furthermore, numerous websites nowadays allow users to rate items based on multiple criteria. In terms of restaurants, users can rate restaurants based on a variety of aspects. On TripAdvisor (https://www.tripadvisor.com/), for example, consumers can rate restaurants based on three criteria: service, food, and value. The multi-criteria ratings of restaurants are used in this study to accurately learn consumers' preferences and, as a result, provide more personalized restaurant recommendations.

Trust-based recommender systems use social networks that are weighted by trust ratings to make recommendations to users based on other users they trust. Trust data can be gathered either explicitly or implicitly. Users' explicit trust information can be collected, and each user can determine whether or not others are trustworthy. Implicit trust information, alternatively, can be inferred from user ratings [15,16]. The implicit trust information is used in this study as a supplementary source of information to alleviate the impact of sparsity and new user drawbacks and, as a result, increase the coverage of recommendations.

To this end, in this paper, we propose a trust-based recommendation model for helping consumers select suitable restaurants in reference to their preferences. Specifically, we propose a Trust-based Multi-Criteria CF (TMCCF) recommendation model that includes multi-criteria item ratings and implicit trust relationships among users for restaurant recommendations. There are a number of advantages of the proposed recommendation model. First, the model exploits the multi-criteria ratings of restaurants to learn the users' preferences more accurately. Second, the model incorporates implicit trust information of users as an additional source of information to overcome the sparsity and new user challenges. This paper is organized as follows. Sections 2 and 3 present related work and introduce the proposed model, respectively. Experimental evaluations and results are demonstrated in Section 4. Finally, we conclude the paper and further work in Section 5.

2. RELATED WORK

Restaurants recommendation is a hot topic among numerous recommendation applications that has attracted the interest of practitioners and researchers in recommender systems in recent years [1-7]. A number of restaurant recommender systems have utilized mobile-based context aware services as well as location-based approaches, for example, Chu and Wu [1] proposed a restaurant recommender system based on mobile context aware services to supply users with personalized restaurant recommendations. The proposed system significantly satisfies the user's needs for restaurant recommendations by utilizing location-based approaches and user preferences. The experimental results of users satisfaction revealed that the proposed system satisfies the search requirements of the mobile users to a great extent. In another study, Zeng et al. [3] developed a restaurant recommendation system based on mobile environment. The proposed system employs a user preference model based on the features of user's previously visited restaurants as well as location information. The Baidu map cloud service and Baidu web cloud service were utilized to find the user's location as well as the nearby restaurants. The results of a case study revealed that the proposed system is capable of successfully recommending appropriate restaurants to a variety of users.

Some restaurant recommendation systems use CF-based approaches to provide suggestions to users. For instance, a restaurant recommendation system, based on a user-based CF approach, was developed by Li et al. [4]. The proposed approach is broken down into three components: user rating similarity, user attributes similarity, and a fusion of these similarities. The experiments were carried out on data from 627 restaurants and 46718 ratings provided by 30081 users from the dianping.com website, which covered the city of Guilin, China. Using MAE and RMSE measurements, the proposed approach is shown to be effective and accurate in recommending restaurants when compared to traditional user-based CF. In similar research, Fakhri et al. [5] propose a user-based CF method for recommending restaurants based on users' ratings and users' attributes. If a target user wants to find a restaurant, then the system will calculate the similarities between the target user and other users based on their ratings and attributes. Then neighbors who have the biggest similarity with the target user will be consulted for personalized restaurants recommendations. For validation purposes, the study uses a questionnaire of users who have rated restaurants they have visited. The dataset contains 86 restaurants from the zomato.com site and 50998 ratings from 593 users from the questionnaire. Furthermore, Tripathi and Sharma [6] propose a restaurant recommender system that employs the k-Nearest Neighbors and the multiclass support vector machine (SVM) classification algorithms. The experiment was carried out using a dataset obtained from the Yelp website. The experimental results reveal that both the user-based SVM and item-based SVM outperform the K-nearest neighbours algorithm.

With the growth in the number of users' reviews on websites and social media, sentiment analysis has become a viable method for extracting users' preferences. As a result, a number of restaurant recommendation systems have been analyzing user reviews to determine their preferences. For example, Alfarhood and Gauch [2] propose Traveltant, a social network-based restaurant recommender system that makes recommendations based on the user's preferences, the preferences of their friends, and the restaurant's overall reputation. To recommend restaurants, Traveltant mines the user's and his/her friends' interests from Facebook and retrieves the restaurants, their categories, and their reputation from Yelp. The proposed system is validated by asking volunteers to rate the recommendation results using 14 distinct models representing different combinations of factors, and the results demonstrate that personal preferences are the most important factor influencing the decision-making process when it comes to where to dine. In another study, a context-aware restaurant recommender system is proposed by Asani et al [7]. The proposed system first applies natural language processing techniques to users' comments about restaurants to extract the desired food names. After that, the names of foods retrieved from user comments are clustered and their sentiments about them are analyzed using a semantic technique. Finally, nearby restaurants that fit the user's food preferences are recommended. The proposed system is evaluated using data from comments on the TripAdvisor website. The evaluation results show that the proposed system can make highly accurate recommendations to users. KesavaDasu et al. [18] proposed a restaurant recommender system using a nearest neighbor based MapReduce approach. The top-k restaurants are retrieved based on the preference of the user's cuisine, the food price, and the distance from the customer's current location. Chen and Xia [19] proposed an approach for restaurant recommendations. In this approach, the user-based CF is integrated with the distance decay function to take into account the geographical effect on the restaurant selection. To provide just-in-time recommendations, the approach filters out restaurants that are not open according to the current time. In order to alleviate the sparsity of data, restaurants are assembled by their price tags. Experiments on the Foursquare dataset demonstrate that the approach outperforms traditional recommendation approaches. Dutta et al. [20] proposed a restaurant recommender system which predicts the rating of new restaurants based on an analysis of ratings, reviews, restaurant type, cuisines, online ordering service, demand, and availability of the restaurant. The authors utilize a RandomForestRegressor to predict the ratings of restaurants. The study uses the Zomato Bengaluru dataset to realize which features are vital to predicting the ratings of new restaurants.

However, in comparison to the massive amount of research done in recent years on other real-world applications of recommender systems, restaurant recommender systems have received less attention.

3. THE TRUST-BASED MULTI-CRITERIA CF (TMCCF) RECOMMENDATION MODEL

The proposed TMCCF model takes as input a raw matrix of user-item MC ratings, which is made up of M users' multi-criteria ratings on N items. The following four primary tasks exemplify the proposed TMCCF recommendation process, as shown in Fig. 1.

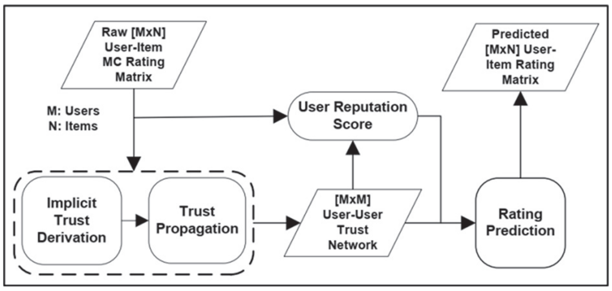

Raw [MxN] User-Item MC Rating Matrix
M: Users
N: Items
User Reputation Score
Predicted [MxN] User-Item Rating Matrix
Implicit Trust Derivation
Trust Propagation
[MxM] User-User Trust Network
Rating Prediction

Fig. 1. The TMCCF recommendation model

3.1 THE DERIVATION OF IMPLICIT TRUST

The reliability of a given user can be obtained by appraising the accuracy of predicted ratings of that user as a recommender to the active user in the past. For this purpose, the predicted rating of user u on item i based on the only neighbor user v is given as follows:

$$P_{u,i} = \overline{r_u} + (U^v(i) - \overline{r_v}), \quad (1)$$

Where $\overline{r_u}$ and $\overline{r_v}$ denote the mean rating of users *u* and *v* respectively. $U^v(i)$ corresponds to the overall utility of user *v* on item *i* defined as follows:

$$U^v(i) = \sum_{a=1}^{k} w^v_a(i) c^v_a(i), \; where \; \sum_{a=1}^{k} w^v_a(i) = 1 \quad (2)$$

Where $c^v_a(i)$ refers to the rating value of user *v* on item *i* in respect of criterion c_a and w^v_a is a weight that implies the importance of criterion c_a by user *v* on item *i*.

Then, based on the distance between the ratings of the co-rated items and predicted ratings, and the importance of the co-rated items, a weighted version of the Euclidean distance [21] with the Inverse User Frequency measure [22] is used to compute the initial implicit similarity of users *u* and *v*.

$$wEuclidean_{u,v} = \frac{1}{1+\sqrt{\sum_{i\in I_{u,v}} \left|P_{u,i} - U^{u}(i)\right|^{2}}} \times Log\left(\frac{|U|}{\left|U_{i\in I_{u,v}}\right|}\right)^{2} \quad (3)$$

Where $P_{u,i}$ is the rating prediction of user *u* on item *i*, and $I_{u,v}$ is co-rated item set of the users *u* and *v* . $U^{u}(i)$ is the overall utility of user *u* on item *i*, $|U|$ is the overall number of users in the rating matrix, and $|U_{i}|$ is the overall number of users who rated item *i*.

The weighted Euclidean distance, on the other hand, still has limitations as users who have rated a small number of items can gain a high level of trust with almost all other users. To solve this problem, we use the Rating Jaccard method [23]. The Rating Jaccard method is a structural similarity metric that considers the proportion of total common ratings that are equivalent in absolute value, calculated as follows:

$$RJacc_{u,v} = \frac{|N_{T(u,v)}|}{|I_{u} \cap I_{v}|} \quad (4)$$

Where $|I_{u} \cap I_{v}|$ is the total number of commonly rated items by users *u* and *v*, and *NT(u,v)* is the total number of commonly rated items that have the same absolute value given as follows:

$$N_{T(u,v)} = \left\{\begin{array}{ll} N_{T(u,v)} + 1; & \text{if } \forall_{\,\mathrm{i}\in I_{u}\cap I_{v}}\ \mathrm{R}_{u,i} = \mathrm{R}_{b,i} \\ N_{T(u,v)} \text{ remains unchanged;} & \text{otherwise} \end{array}\right\} \quad (5)$$

In addition, to address the sparsity issue, an extreme behavior similarity metric [24] is used as a weighting factor. The extreme behavior similarity suggests that users who give an extreme rating (like 1 or 5) on the same item are more similar than users who give a neutral rating (like 3), and a user's exceptional rating on an item is more significant than the public rating.

$$ExBSim_{u,v} = \frac{\sum_{i\in I_{u,v}} S1(u_i, v_i) \times S2(u_i, v_i)}{\sqrt{\sum_{i\in I_{u,v}} S1^{2}(u_i, v_i)} \times \sqrt{\sum_{i\in I_{u,v}} S2^{2}(u_i, v_i)}} \quad (6)$$

Where *S1(u$_i$,v$_i$)* reflects the influence of users *u* and *v's* extreme ratings on item *i* compared to the median rating on the system, defined as follows:

$$S1(u_i, v_i) = \frac{1}{1+\exp(-\left|r_{u,i} - \bar{r}_{med}\right|\left|r_{v,i} - \bar{r}_{med}\right|} \quad (7)$$

Where *S2(u$_i$,v$_i$)* reflects the influence of users *u* and *v's* extreme ratings on item *i* compared to its mean rating, defined as follows:

$$S2(u_i, v_i) = \frac{1}{1+\exp(-\left|r_{u,i} - \bar{r}_{i}\right|\left|r_{v,i} - \bar{r}_{i}\right|} \quad (8)$$

Finally, the implicit trust between any given pair of users is calculated as follows:

$$iTrust_{u,v} = wEuclidean_{u,v} \times RJacc_{u,v} \times ExBSim_{u,v} \quad (9)$$

3.2 THE COMPUTATION OF TRUST PROPAGATION

Trust propagation is the notion that contributes to the formation of new trust relationships from pre-existing trust relationships. Trust transitivity is the most visible form of trust propagation, which means that if *X* trusts *Y*, and *Y* trusts *Z*, *X* will likewise trust *Z* due to transitivity. In this study, we employ trust propagation to compute the implicit trust among users who do not have direct relationships in the implicit trust social network. To compute the propagated implicit trust between users, we use the following aggregation metric.

$$pTrust_{u,h} = \frac{\sum_{v\in adj(u\,and\,h)} \left(iTrust_{u,v} \times URJacc_{u,v}\right) + \left(iTrust_{v,h} \times URJacc_{v,h}\right)}{\sum_{v\in adj(u\,and\,h)} RJacc_{u,v} + RJacc_{v,h}} \quad (10)$$

Where *adj(u and h)* is the set of trusted adjacent neighbors of user *u* who trust user *h*, which includes user *v*.

3.3 THE COMPUTATION OF USER REPUTATION SCORE

The user reputation score is used to improve the system's capacity to predict unseen items that are caused by an active user's lack of trusted nearest neighbors. It's determined based on the average variation between the user items rating and the items' mean rating, as well as the ratio of trust relationships with other users in the implicit trust social network [25], as shown below.

$$URS_{u} = \exp\left(-\frac{\sum_{i\in I_{u}} \left|r_{u,i} - \bar{r}_{i}\right|}{|I_{u}|}\right) \times \sqrt{\frac{|U_{u}|}{|U|}} \quad (11)$$

Where $r_{u,i}$ is the rating of user u on item *i*, $\bar{r}_{i}$ is the mean rating of item *i* by all users, and $|U_{u}|$ is the total number of users who are associated to user *u* in the implicit trust social network.

3.4 THE COMPUTATION OF RATING PREDICTION

For the computation of the final predicted rating, the deviation-from-mean metric is used, as shown below:

$$P_{u,i} = \left\{\begin{array}{ll} \bar{r}_{u} + \frac{\sum_{v\in N^{U}} Trust_{u,v} \times (r_{v,i} - \bar{r}_{v})}{\sum_{b\in N^{U}} Trust_{u,v}}; & \text{if } Trust_{u,v} \neq 0 \\ \bar{r}_{u} + \frac{\sum_{v\in N^{U}} URS_{v} \times (r_{v,i} - \bar{r}_{v})}{\sum_{v\in N^{U}} URS_{v}}\ ; & \text{if } Trust_{u,v} = 0 \end{array}\right. \quad (12)$$

Where $Trust_{u,v}$ represents the trust obtained from the implicit trust social network between the user *u* and neighbour user *v*, $r_{v,i}$ is the rating of item *i* by user *v*. URS_{v} is the user's reputation score of user *v*, and N^{U} is the nearest neighbour set of user *u*.

4. PERFORMANCE EVALUATION

4.1 DATASET AND EVALUATION MEASURES

The Restaurant MC dataset gathered from the TripAdvisor website is used to validate the performance of the proposed TMCCF recommendation model. The dataset comprises users' numerical ratings of restaurants on three criteria: food, service, and value. The rating values range from 1 to 5. It contains 14,633 multi-criteria ratings for 205 restaurants from 1,254 users. The dataset is split into 80% training and 20% test sets for the performance evaluation.

To measure the effectiveness of the proposed model, the recommendations made were assessed using two metrics: 1) the Mean Absolute Error (MAE) metric, which measures how much the predicted ratings are close to the actual ratings (the smaller the value of MAE, the more accurate a recommendation), and 2) the prediction Coverage metric, which is the proportion of prediction requests for which a recommendation algorithm can produce a prediction [26].

4.2 METHODS FOR COMPARISON

The results of the proposed TMCCF recommendation model were compared to the results of two commonly used benchmark user-based CF recommendation algorithms: 1) the user-based SC CF algorithm [27] that employs Pearson Correlation as a similarity measure to produce personalized recommendations (denoted by SC-UCF); and 2) the user-based MC CF algorithm [21], which adopts the similarity-based approach to produce personalized recommendations (denoted by MC-UCF).

4.3 PERFORMANCE COMPARISON

Extensive experiments have been conducted to assess the performance of the proposed TMCCF recommendation model, respecting the prediction accuracy and prediction coverage when confronted with the limitations of sparsity and new user.

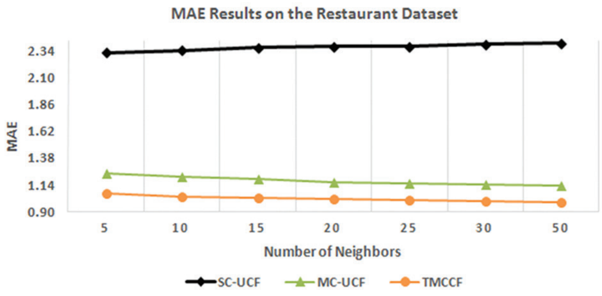


Fig. 2. MAE performance under varying numbers of neighbours on the Restauarent dataset

Performance evaluation on the Restaurant dataset. In comparison to the other benchmark algorithms on the Restaurant dataset, the proposed TMCCF model produces the best results in terms of the MAE measure under varying numbers of nearest neighbors, as shown in Fig 2. The MAE improvement results of the proposed model are roughly 57% and 13% better than the benchmark algorithms, respectively. Notably, the results illustrate that the proposed model outperforms the benchmark algorithms on the Restaurant dataset in terms of prediction accuracy.

Performance evaluation with varying Sparsity levels. A number of experiments were performed on different datasets with varying sparsity levels by filtering out users who provided ratings less than a specific number of times in the given datasets. The experimental results of the proposed TMCCF and benchmark algorithms are shown in Fig. 3 and Fig. 4. We can notice that TMCCF consistently outperforms the other benchmark algorithms in terms of MAE and Coverage in all cases. The TMCCF model obtains lower MAE values and higher Coverage percentages than the benchmark algorithms at each sparsity level, which further demonstrates that the incorporation of implicit trust information of users is very helpful in finding adequate neighbours, leading to increased prediction accuracy and coverage in highly sparse datasets.

The results demonstrate that the proposed model's MAE is improved by roughly 69% and 66%, respectively, when compared to the benchmark algorithms. Coverage has increased by around 56% and 49%, respectively. Incredibly, the proposed model outperforms benchmark algorithms in dealing with highly sparse datasets, as evidenced by the considerable improvement in MAE and Coverage results.

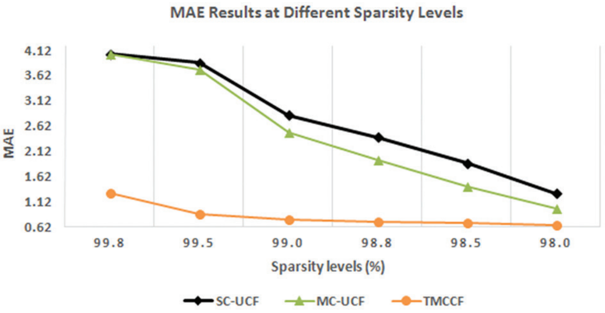


Fig. 3. MAE performance under varying levels of sparsity

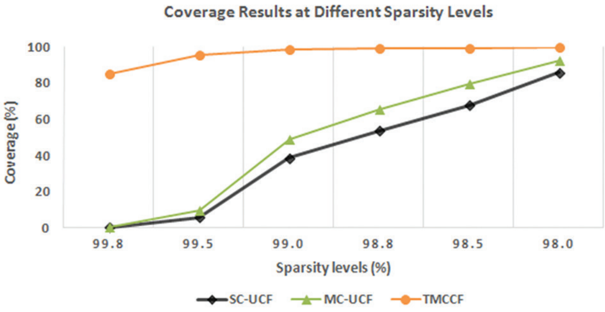


Fig. 4. Prediction coverage performance under varying levels of sparsity

Performance evaluation with a varying number of ratings of new user. A number of experiments were conducted on six datasets with a varying number of ratings of a new user (from 10 to 20 ratings) in the given datasets. Fig. 5 and Fig. 6 demonstrate the experimental results of the proposed TMCCF and benchmark algorithms. It can be noticed that TMCCF constantly exceeds the other benchmark algorithms in terms of MAE and Coverage in all cases. At each number of ratings, the TMCCF model achieves lower MAE values and higher Coverage percentages than the benchmark algorithms, which again confirms that including implicit trust information of users is very helpful in finding sufficient neighbours, resulting in increased prediction accuracy and coverage in new user situations.

When compared to the benchmark algorithms, the results demonstrate that the proposed model's MAE is improved by around 74% and 69%, respectively. Coverage has increased by around 65% and 51%, respectively. As indicated by the significant improvement in MAE and Coverage results, the proposed model outperforms benchmark algorithms in reducing the impact of the new user problem.

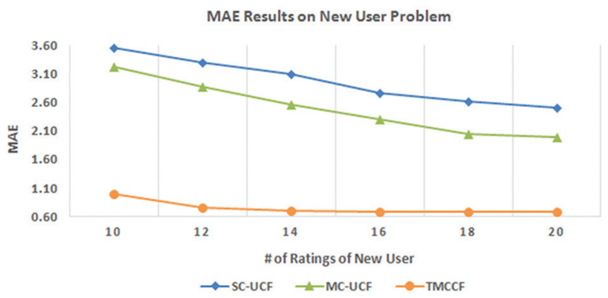


Fig. 5. MAE performance under a varying number of ratings of new users

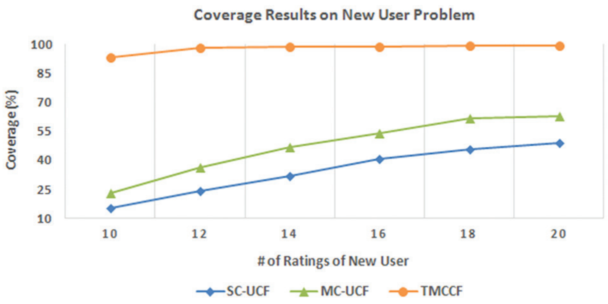


Fig. 6. Prediction coverage performance under a varying number of ratings for new users

5. CONCLUSION AND FUTURE WORK

This study proposes a trust-based recommendation model to assist consumers in selecting appropriate restaurants based on their preferences. For high-quality personalized restaurant recommendations, the proposed model utilizes multi-criteria ratings and implicit trust relationships among consumers. The proposed recommendation model has several advantages. First, the model makes use of multi-criteria restaurant ratings to better understand the consumers' preferences, resulting in improved prediction accuracy. Second, to compensate for the lack of ratings, the model exploits implicit trust information from users as an additional source of information, which improves prediction coverage by reducing the impact of sparsity and new user problems. The proposed model achieves significantly better performance as compared to benchmark CF-based recommendation algorithms in both prediction accuracy and prediction coverage. The model demonstrates its significance in overcoming the poor prediction accuracy and coverage caused by sparsity and new user issues. In the future, we will focus on analyzing the impact of incorporating other resources of information into the recommendation process, such as users' reviews of restaurants.

6. REFERENCES:

[1] C. Chu, S. Wu, "A Chinese Restaurant Recommendation System Based on Mobile Context-Aware Services", Proceedings of the 14th International Conference on Mobile Data Management, Milan, Italy, 3-6 June 2013, pp. 116-118.

[2] S. Alfarhood, S. Gauch, "Traveltant: Social Network-Based Restaurant Recommender System", Proceedings of the 14th IADIS International Conference on WWW/Internet, Maynooth, Greater Dublin, Ireland, 24-26 October 2015, pp. 47-54.

[3] J. Zeng, F. Li, H. Liu, J. Wen, S. Hirokawa, "A Restaurant Recommender System Based on User Preference and Location in Mobile Environment", Proceedings of the 5th IIAI International Congress on Advanced Applied Informatics, Kumamoto, Japan, 10-14 July 2016, pp. 55-60.

[4] L. Li, Y. Zhou, H. Xiong, C. Hu, X. Wei, "Collaborative filtering based on user attributes and user ratings for restaurant recommendation", Proceedings of the 2nd Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, 25-26 March 2017, pp. 2592-2597.

[5] A. A. Fakhri, Z. Baizal, E. B. Setiawan, "Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region", Proceedings of the 2nd International Conference on Data and Information Science, Bandung, Indonesia, 2019, pp. 1-7.

[6] A. Tripathi, A. K. Sharma, "Recommending Restaurants: A Collaborative Filtering Approach", Proceedings of the 8th International Conference on Reliability, Infocom Technologies and Optimization, Noida, India, 4-5 June 2020, pp. 1165-1169.

[7] E. Asani, H. Vahdat-Nejad, J. Sadri, “Restaurant recommender system based on sentiment analysis”, Machine Learning with Applications, Vol. 6, No. 2021, 2021, pp. 1-10.

[8] Q. Shambour, A. H. Hussein, M. Abualhaj, Q. Kharma, “An Effective Hybrid Content-based Collaborative Filtering Approach for Requirements Engineering”, Computer Systems Science and Engineering, Vol. 40, No. 1, 2022, pp. 113-125.

[9] Q. Y. Shambour, “Hybrid recommender systems for personalized government-to-business e-services”, University of Technology Sydney, Faculty of Engineering and Information Technology, Sydney, Australia, Phd Thesis, 2012.

[10] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, “Recommender system application developments: A survey”, Decision Support Systems, Vol. 74, 2015, pp. 12-32.

[11] Q. Shambour, S. Fraihat, M. Hourani, “The Implementation of Mobile Technologies in Higher Education: A Mobile Application for University Course Advising”, Journal of Internet Technology, Vol. 19, No. 5, 2018, pp. 1327-1337.

[12] Q. Shambour, J. Lu, “An effective recommender system by unifying user and item trust information for B2B applications”, Journal of Computer and System Sciences, Vol. 81, No. 7, 2015, pp. 1110-1126.

[13] S. Fraihat, Q. Shambour, “A Framework of Semantic Recommender System for e-Learning”, Journal of Software, Vol. 10, No. 3, 2015, pp. 317-330.

[14] G. Adomavicius, A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions”, IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6, 2005, pp. 734-749.

[15] Q. Shambour, N. Turab, O. Adwan, “An Effective e-Commerce Recommender System Based on Trust and Semantic Information”, Cybernetics and Information Technologies, Vol. 21, No. 1, 2021, pp. 103-118.

[16] Q. Shambour, “A user-based multi-criteria recommendation approach for personalized recommendations”, International Journal of Computer Science and Information Security, Vol. 14, No. 12, 2016, pp. 657-663.

[17] A. Hussein, Q. Kharma, F. Taweel, M. Abualhaj, Q. Shambour, “A Hybrid Multi-Criteria Collaborative Filtering Model for Effective Personalized Recommendations”, Intelligent Automation & Soft Computing, Vol. 31, No. 1, 2022, pp. 661-675.

[18] K. KesavaDasu, K. R. Kamal, G. Pranith, “A Cuisine Based Recommender System Using k-NN and Mapreduce Approach”, International Journal of Recent Advances in Multidisciplinary Research, Vol. 1, No. 1, 2021, pp. 9-17.

[19] L. Chen, M. Xia, “A restaurant recommendation approach with the contextual information”, Journal of Intelligent & Fuzzy Systems, Vol. 40, No. 3, 2021, pp. 4481-4489.

[20] K. B. Dutta, A. Sahu, B. Sharma, S. S. Rautaray, M. Pandey, "Machine Learning-Based Prototype for Restaurant Rating Prediction and Cuisine Selection", Advances in Intelligent Systems and Computing, Singapore, Springer, 2021, pp. 57-68.

[21] G. Adomavicius, Y. O. Kwon, “New recommendation techniques for multicriteria rating systems”, IEEE Intelligent Systems, Vol. 22, No. 3, 2007, pp. 48-55.

[22] J. S. Breese, D. Heckerman, C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", Proceedings of the Fourteenth conference on Uncertainty in Artificial Intelligence, Madison, WI, USA, 1998, pp. 43–52.

[23] M. Ayub, M. A. Ghazanfar, T. Khan, A. Saleem, “An Effective Model for Jaccard Coefficient to Increase the Performance of Collaborative Filtering”, Arabian Journal for Science & Engineering, Vol. 45, No. 12, 2020, pp. 1-21.

[24] C. Feng, J. Liang, P. Song, Z. Wang, “A Fusion Collaborative Filtering Method for Sparse Data in Recommender Systems”, Information Sciences, Vol. 521, 2020, pp. 365-379.

[25] H. Song, Q. Pei, Y. Xiao, Z. Li, Y. Wang, "A Novel Recommendation Model Based on Trust Relations and Item Ratings in Social Networks", Proceedings of the International Conference on Networking and Network Applications, Kathmandu, Nepal, 16-19 October 2017, pp. 17-23.

[26] C. C. Aggarwal, "Evaluating Recommender Systems", Recommender Systems: The Textbook, Springer International Publishing, 2016, pp. 225-254.

[27] J. Herlocker, J. A. Konstan, J. Riedl, “An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms”, Information Retrieval, Vol. 5, No. 4, 2002, pp. 287-310.

Bolstering user authentication: a kernel-based fuzzy-clustering model for typing dynamics

Original Scientific Paper

Anthony Metumaraibe Nwohiri
University of Lagos,
Faculty of Science, Department of Computer Sciences
University Road, Akoka-Yaba, Lagos Nigeria
anwohiri@unilag.edu.ng

Ufuoma Cyril Ogude
University of Lagos,
Faculty of Science, Department of Computer Sciences
University Road, Akoka-Yaba, Lagos Nigeria
uogude@unilag.edu.ng

Hai Vinh Nguyen
Vietnam National University,
Faculty of Mathematics, Mechanics and Informatics,
Department of Informatics
Nguyen Trai Rd, Hanoi, Vietnam
nguyenhaivinh@vnu.edu.vn

Edilberto Moreno Sanchez
Autonomous National University of Mexico,
Astronomy Institute, Department of Information Technology and Scientific Computing
Coyoacán borough, Mexico City, Mexico
edilberto@astro.unam.mx

Abstract – *In most information systems today, static user authentication is accomplished when the user provides a credential (for example, user ID and the matching password). However, passwords appear to be the most insecure authentication method as they are vulnerable to attacks chiefly caused by poor password hygiene. We contend that an additional, non-intrusive level of security can be achieved by analyzing keystroke biometrics and coming up with a unique biometric template of a user's typing pattern. The paper proposes a new model for representing raw keystroke data collected when analyzing typing biometrics. The model is based on fuzzy sets and kernel functions. The corresponding algorithm is developed. In the static authentication problem, our model demonstrated relatively higher performance than some classic anomaly-detection algorithms, such as Mahalanobis, Manhattan, nearest neighbor, outlier counting, neural network, and the support-vector machine.*

Keywords: *anomaly detection, data mining, fuzzy clustering, keystroke biometrics, kernel function, machine learning, static authentication*

1. INTRODUCTION

Internet proliferation has exploded over the past decade. This has made efficient access control (AC) for information systems ever more challenging. Intruders may gain access to systems from virtually anywhere over the Internet. As a fundamental concept in security, AC regulates who or what can view or use resources in a computing environment. It restricts access to computers, networks, applications, files and other sensitive data. One of the crucial functions of AC systems is to ensure that "someone" is who they claim to be (authentication) and that they have the appropriate data access (authorization). Hence, user authentication is a major important challenge.

The user authentication problem can be tackled in many ways. In modern information systems, passwords remain the most common digital authentication method [1–4]. Passwords are typically a string of characters used to confirm a user's identity during the authentication process. While passwords are a weak form of protection, their simplicity makes them easy to use and administer. However, passwords are among the most vulnerable authentication methods. Poor password hygiene is a top cause of data breaches. A dictionary attack can be used to pick up the passphrase. Password hash can be leaked directly from their storage location to be cracked offline [4]. Digital signature systems (DSSs), also used for authentication, are free from dictionary attacks [5]. However, DSSs cannot guarantee secure storage of private keys because the keys are stored using other access control means [6, 7]. Moreover, digital signature needs to be verified and there is no legal backup for this verification process [6].

In view of the above, most high-security systems use biometrics to identify and authenticate individuals. The primary premise of biometric authentication is that any user can be precisely identified by intrinsic physical or behavioral traits. This authentication method comes with several benefits – it is convenient, every user has access to a unique set of biometrics, biometrics are hard to steal, and of course this technique comes with high security and assurance.

Biometric authentication uses unique biological traits to verify that someone is who they say they are. Such traits include palm print, fingerprints, hand geometry, voice, retinal patterns, iris recognition, facial recognition, DNA, odor/scent, and hand patterns [8]. Biometric systems can be divided into two categories: (a) physical biometrics systems – are effective, but quite expensive as they require special hardware; (b) behavioral biometrics systems – they analyze parameters such as a user's keystrokes dynamics, navigational patterns, screen pressure, typing speed, mouse or mobile movements, gyroscope position and more [9-11]. They do not require any special hardware and are easy to implement.

In this paper we first present the problem at hand, and look at existing user authentication methods. After that, we propose a keystroke data representation model and develop the corresponding algorithm. The algorithm is compared with some classic anomaly detectors.

2. PROBLEM STATEMENT

This research deals with the static authentication problem. Static authentication reuses a specific authenticator (e.g., static password). This type of authentication only provides protection against attacks in which an imposter cannot obtain the authenticator. An authentication process is strong if it is difficult to guess or decrypt the authenticator values and if the values themselves are secured in transit and while stored on the system [12].

The method works on a known pattern or other predefined text. The data (e.g., password) entered by the user when attempting to login is collected and compared with previous successful login attempts. This technique is an extension of the standard user ID/password-based authentication method (i.e., the system checks not only what the user typed, but how it was typed).

Several static authentication features are worth noting. First, the input data is relatively small. Typically, static authentication works in tandem with password authentication. This practically eliminates the use of extremely long passwords (over 100 characters) that the user would have to manually type.

Another feature is that input data is static in nature. The password, for the most part, remains the same from the moment it is created till it is changed or updated. In this case, only a small amount of typing biometrics data can be extracted. Besides, more often than not, the password is changed very rarely, which leads to a large sample. Therefore, static authentication should be optimized to be able to recognize a user based on a small set of parameters.

Static authentication should be completed as quickly as possible since the user will not be granted access until authentication data has been successfully processed. Until authentication is completed, the user will not be allowed to access the system. Therefore, the time interval between the moment the user enters his password and the moment he gains access to the system should be minimized as much as possible.

So, to put it more formally, our authentication problem can be described as follows. We assume there is a set of users performing certain keyboard actions, for example, pressing a key or typing a password. Here, the training task is to match a certain function (model) to each user; the model is to serve as a measure of the anomalousness of user actions, i.e., it will be used to verify the identity of real users and detect anomalous users. The authentication task is to measure (based on the model) the anomalousness of a new user action and process the calculated value. This value determines whether user authentication would be accepted or rejected.

The false rejection rate (FRR) and false acceptance rate (FAR) were used to evaluate obtained results. False rejection occurs when an authentic user is rejected by the system, while false acceptance is when an imposter is accepted. A lower FRR means less rejection and easier access by genuine users. A lower FAR indicates less imposter accepted. [13-15].

Authentication algorithms, including the one proposed in this paper, do not return a binary value – authentication successful/authentication failed. Rather, they return some real value indicating how well the authentication attempt matched the training data. Hence, it is necessary to introduce a certain threshold that would distinguish between accepted and rejected authentication attempts. Varying this real value, we can find a threshold at which FRR and FAR are equal. That is, the authentication system is configured such that the rate of false negatives and the rate of false positives are approximately equal. The crossover error rate (CER) describes the point where FRR and FAR are equal. It is one of the most important indicators used in evaluating the performance of any biometric security system [16, 17]. The CER describes the overall accuracy of a biometric system. Figure 1 shows that the lower the CER value, the higher the accuracy of the biometric system. If the threshold of acceptance (sensitivity) of the system is increased, FRR will increase and FAR will fall. Likewise, choosing a low threshold will result in high FAR and low FRR [15, 18].

Since the data used for the experiment contains password typing biometrics from many people, it can be assumed that the algorithm's performance will vary depending on the subject. Obviously, a highly efficient algorithm should produce equally good results regardless of the subject being authenticated. The greater the CER spread for different subjects, the harder it would be using the algorithm in practice. Therefore, the standard deviation of CERs for different subjects becomes another important parameter for evaluating the algorithm's performance.

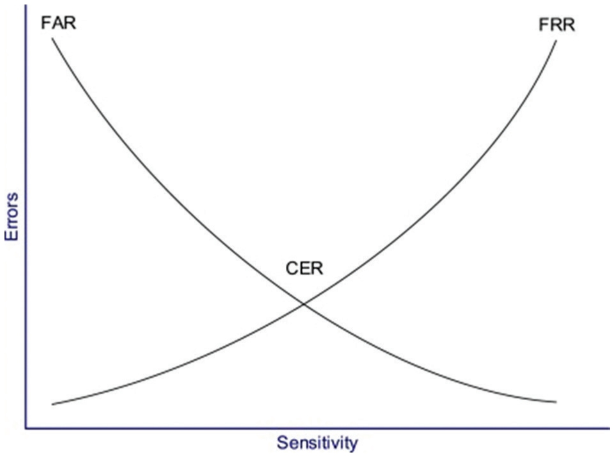


Fig. 1. False acceptance rate versus false rejection rate

3. EXISTING AUTHENTICATION APPROACHES

Existing authentication methods are based on various features collected while the user is typing at a computer keyboard. They are also based on the models (used to verify the real user's identity) created using these features.

3.1. DWELL TIME

Dwell time is the period during which a key is in a pressed state. As described by Wong et al. [19], the key hold method takes a vector as a model, the vector consists of some elements. Each of the elements correspond to a key on the keyboard, representing a pair – the mean dwell time of the key and the standard deviation for that key. So, the dwell time of a key (user action) is considered abnormal if the difference between it and the mean dwell time for that key is greater than the standard deviation for that key. A percentage of allowed abnormal actions is given. If this threshold is exceeded, authentication will be rejected. The authors achieved FRR <10% and FRR <10% all at a time.

3.2. KEY EVENT ORDER

Lau et al., [20] observed that when typing at the keyboard, some people sometimes unconsciously press a second key before releasing the first one. The authors called this phenomenon a "swap". A user model can be constructed by observing the key-press/key-release sequence and considering the number of "swaps". To obtain an anomaly score, the distance between the tested and the user set was compared with the average distance between the keystrokes of the same user. An authorization attempt was rejected whenever the distance went beyond one standard deviation. The resulting FAR and FRR depended strongly on pairs of users, with error rates ranging from 0% to 70%. This clearly makes the method unsuitable for use unless with other approaches.

3.3. RELATIVE TYPING SPEED

It is assumed that for each pair of keys, the typing speed remains about the same regardless of the text being typed. Therefore, it was suggested that the typing speeds of pairs of keys be measured and used as a user model. The distance between vectors of key pairs, ordered by typing speed, as proposed by Bergadano et al. [21], was used to build a model. Based on the resulting mean distance between any two vectors of the same user and between any two vectors of different users, the researchers suggested that an authentication attempt should be accepted only when the difference in vectors is less than 0.33 and rejected when it is greater than 0.66.

3.4. THE *SHIFT* KEYS

In their paper, Lau et al. [20] argued that people use the right and left *Shift* keys differently, and that this could be used for authentication. The authors divided users into 4 groups: strictly left-*Shift* users, strictly right-*Shift* users, those who used the left *Shift* key more often than the right, and those who used the right *Shift* key more often than the left. Obviously, if a user hits the expected group, that does not necessarily mean that any authentication attempt made by that user must be accepted since we have a limited number of groups (only 4), and the false acceptance rate is very high. However, when a user hits the wrong group, that authentication attempt will be rejected.

3.5. SHORT ALPHABETIC OR NUMERIC PASSWORDS

Techniques proposed in [22] and [23] used keystroke times as a model. However, in [22], times were measured using three different typologies, as shown in Table 1.

Table 1. Typologies of time.

Time topologies	Description
Absolute time	Consists of the dwell time (how long a key was held pressed) and the flight time (the duration between the moment the key was released and the moment the next one was pressed)
Cumulative time	Consists of the accumulated absolute times for typing a particular phrase. This allows to smooth out outliers.
Ratio time	This is the ratio of the dwell time to the flight time

A multiclass linear support vector machine (SVM) was used as the training algorithm, as it demonstrates high results on simple-structure data [24].

During data collection, the subjects were divided into two groups: those who were informed about the experiment and those who were not. It was demonstrated that obtained results were a function of users' awareness of the experiment. Specifically, the obtained FAR and FRR (1%-3%) were 3-5 times lower among the informed users.

4. PROPOSED APPROACH

4.1. DESCRIPTION OF THE MODEL AND ALGORITHM USED

Keystroke time intervals are the main inputs of the model. The various time intervals used for the feature space of our algorithm are presented in Table 2. The feature space was chosen as follows. Each password typing attempt is represented as a vector of time lapses between different key-press and key-release events. A key-press event occurs when a key that produces a character value is pressed down, while a key-release event occurs once the key is released.

Keystroke time intervals are the main inputs of the model. The various time intervals used for the feature space of our algorithm are presented in Table 2. The feature space was chosen as follows. Each password typing attempt is represented as a vector of time lapses between different *key-press* and *key-release* events. A *key-press* event occurs when a key that produces a character value is pressed down, while a *key-release* event occurs once the key is released.

Table 2. Timing vector for password-input events.

Time intervals	Description
Dwell time	The period, during which a key is in a pressed state. In other words, it is the length of time a key is pressed until it is released.
Press–press	Interval between two successive key presses (always positive)
Release–press	Interval between a key release and the next key press time (may be negative if next key is pressed before previous key is released)
Release–release	Interval between two successive key releases (always positive)

It should be noted that for the first key pressed, only the dwell time is measured.

We used the outlier detection technique presented in [25] as the static authentication problem. The method is based on two main ideas: using kernel functions to define distances and deploying the fuzzy set theory to build a user model. We have adapted the approach to our problem.

4.1.1 KERNEL FUNCTIONS

Kernel functions (KFs) provide a way to manipulate data. The function of kernel is to take data as input and transform it into the required form. Kernels represent a method of computing the dot product of two vectors in a certain feature space. They are widely used in various machine learning algorithms [26-30] and have been shown to be very efficient in tackling various attack/intrusion detection problems [31-33]. KFs allow for efficiency in biometric security systems. These functions enable you to avoid the trouble of having to go into an infinite-dimensional space; they also save you the time that would have been spent on computing map functions.

Nonlinear mapping from an input space of objects to the feature space is central in kernel methods. The kernel trick is a simple method which involves performing the mapping and the inner product simultaneously by defining its associated KF. The KF, see (1), computes, and returns the inner product between two inputs in the feature dimension.

$$K(x,y) = f(x) \cdot f(y) \tag{1}$$

Here *K* is the kernel function, *x* and *y* are *n*-dimensional inputs, *f* is a feature map from *n*-dimension to m-dimension space, *x·y* denotes a dot product. Usually, *m* is much larger than *n*. The Hilbert space serves as our *m*-dimension space.

Kernel function *K(x,y)* measures the distance (similarity) between two input objects *x* and *y*. This metric can be used to build distance functions. Kernels provide a way of computing dot products in some feature space without even knowing what this space is and what the map *f* is. So, there is no need to compute *f(x)* and *f(y)*; moreover, mapping is implicitly determined by *K(x,y)*. Consequently, computing time and memory costs is also not needed. This is where the basic advantages of kernel functions come in.

While classical kernel-based clustering algorithms are based on a single kernel, in practice it is often desirable to base clustering on combination of multiple kernels [34]. The use of different kernels adds a certain flexibility to our approach. It also expands possible configurations for the method, which can be selected such that optimal results are achieved.

In our approach, the following distance function based on the kernel function is considered:

$$d(x,y) = \sqrt{K(x,x) - 2K(x,y) + K(y,y)} \cdot \tag{2}$$

We will focus more on the use of dot products (see expression 1) as kernels and the use of the Gaussian kernel (see expression 3).

$$K(x,y) = \frac{e^{-(x-y)^2}}{2\sigma^2} \tag{3}$$

In expression (3), sigma σ is the standard deviation of the Gaussian distribution. It basically controls how "fat" the kernel function is going to be. It controls the variance around a mean value of the Gaussian distribution (how closely the values of a data set are clustered around the mean). As σ becomes larger, the more variance (allowed around mean) can be chosen to achieve the best results. Conversely, as σ becomes smaller, the less variance allowed around mean can be chosen to achieve the best results. The Gaussian kernel transforms the dot product in the infinite dimensional space into a Gaussian function of the distance between points in the data space.

4.1.2 FUZZY CLUSTERING IN FEATURE SPACE

Introduced independently by Lotfi A. Zadeh and Dieter Klaua in 1965 [35, 36], fuzzy sets were an extension of the classical notion of set [37]. They are objects with a continuum of grades of membership. These sets are characterized by a membership function that maps from the universal set to a value between 0 and 1.

In our proposed method, we search for one common fuzzy cluster containing images of all objects from the original space *X*. **T**he degree of membership of the image of an object from *X* quantifies the grade of membership of that object to each fuzzy cluster, i.e., a value inverse to the anomaly. Images with "small" membership grades (less than a threshold established for a user) will be considered as illegitimate authentication attempts.

Petrovsky [25] demonstrated that a search for a fuzzy cluster in a feature Hilbert space, as we suggested above, results into the following fuzzy clustering problem

$$\left.\begin{array}{c}\min\limits_{\boldsymbol{U\in[0,1]^N,\ a\in H}} \boldsymbol{J(D,c)}\ , \\ (D,c) = \sum_{i=1}^{N}(D_i)^m(f(x_i)-c)^2 - \eta\sum_{i=1}^{N}(1-d_i)^m\end{array}\right\} \tag{4}$$

where *H* is the feature space containing vectors representing authorization attempts; c is the center of the fuzzy cluster in the feature space corresponding to legitimate user authorization attempts; *D* is the function of the membership degree vector; *N* is the number of legitimate authorization attempts used for training; d_i∈[0,1] is the membership degree of image $f(x_i)$ with respect to the fuzzy cluster in the feature space, and, accordingly, the typicalness degree of object xi; m is the fuzziness degree, and eta (*η*) is the distance from the cluster center, where the typicalness degree of the object is considered to be 0.5.

In our method, unlike in some classic fuzzy clustering methods, the center of the cluster or the values $f(x_i)$ cannot be clearly expressed. Nonetheless, *J(D,c)* can be minimized using the following iteration algorithm based on randomized block-coordinate descent [38–40].

First, *D* and *η* are initialized: two points in the training set at maximum distance from each other are found; *η* is taken equal to the square of the distance between these two points and does not change throughout the algorithm; elements of D are taken equal to each other, d_i^0=1/*N*; that is, the cluster center is the same with the "center of gravity" of the images of points *x*. For such a selection of *η*, the typicalness degree of the objects from the learning set is always greater than 0.5.

Second, the cluster center is then calculated.

$$c \cdot c = \left(\sum_{j=1}^{N} d_i^m \sum_{i=1}^{N} d_i^m K(x_i, x_j)\right) \Big/ \left(\sum_{i=1}^{N} d_i^m\right)^2 \tag{5}$$

Third, the distance to the new cluster center is computed for all *j*∈[1,*N*].

$$f(x_j) \cdot c = \left(\sum_{i=1}^{N} d_i^m K(x_i, x_j)\right) \Big/ \left(\sum_{i=1}^{N} d_i^m\right) \tag{6}$$

Fourth, new degrees of membership of training vectors are computed for all *j*∈[1,N]:

$$d_j = \frac{1}{1 + \left(\dfrac{c \cdot c + K(x_i, x_j) - 2\big(f(x_j) \cdot c\big)}{\eta}\right)^{m-1}} \tag{7}$$

The second, third and fourth steps are repeated until:

$$\|D^l - D^{l-1}\| < \varepsilon \tag{8}$$

where *l* is the step number, *ε* is the required accuracy.

In this case, anomaly function *F(x,X)*, which calculates the typicalness degree of a new object takes the form.

$$F(x, X) = \frac{1}{1 + \left(\dfrac{c \cdot c + K(x, x) - 2(f(x, X) \cdot c)}{\eta}\right)^{m-1}} \tag{9}$$

where

$$\begin{gathered}\boldsymbol{f(x,X) \cdot c} = \left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{d_i^m K(x_i, x)}\right) \Big/ \left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{d_i^m}\right) \\ \boldsymbol{d_1 \dots d_N \in X, |X| = N}\P\end{gathered} \tag{10}$$

In this method, *η* was chosen as the square of the distance between two points at maximum distance from each other in the training set. It does not change throughout the algorithm. We call this a simplified variant.

However, there is a more complex way of choosing *η*. The square of the distance from the center of the cluster to the farthest non-outlier vector is used to estimate the cluster radius at each iteration. Outliers are suggested to be the fraction of vectors farthest from the cluster center, which is a parameter of the algorithm. The degree of membership of an object image to the fuzzy cluster in the feature space may be viewed as a typicalness degree of the object. In this case, typicalness degree *F(x,X)* will be > 0.5 if *x* lies inside the cluster, < 0.5 if it lies outside the cluster, or equal to 0.5 if it lies on the border of the cluster. Therefore, in implementing the model, 0.5 is used as the initial minimum typicalness degree by which a typing attempt is to be considered legitimate.

The simplified variant comes with some merits. Its basic advantage is that the anomaly function is continuous, which makes it possible to compare typicalness degrees of objects and also to modify the outlier factor criterion without reconstructing the model. Moreover, the proposed algorithm is considerably simpler than those used for solving quadratic programming problems [25]. The complexity of the algorithm itself is linear.

However, calculation of the kernel matrix has O(n^2) time complexity, where n is the size of the learning set [41].

4.1.3 DATA PRE-PROCESSING

Features in the collected data may be heterogeneous, and their values may have different bounds. Therefore, given the peculiarities of the suggested algorithm, the data should be normalized, bringing the range of values to common boundaries for all the features. The best normalization method for this problem was chosen experimentally on a standard dataset – normalization to the absolute deviation value. Suppose that a feature *p* is encountered in training for *N* password-typing attempts. Then for *p*, the normalization factor for vector *x* would be:

$$\boldsymbol{W_p} = \sum_{i=1}^{N} \frac{|\boldsymbol{x_i} - \overline{\boldsymbol{x}}|}{\boldsymbol{N}}, \ \boldsymbol{x'_p} = \frac{\boldsymbol{x_p}}{\boldsymbol{W_p}} \tag{11}$$

where $\overline{\boldsymbol{x}}$ is the arithmetic mean of the elements of *x*, and *x*' is the normalized vector of *p*.

Among other possible normalization factors that can be used are the square root of the above value, the interquartile range, and some other factors. However, as would be shown later, absolute deviation gives the best results.

5. EXPERIMENT

We conducted a series of experiments in order to compare the suggested method with existing ones and select optimal classification parameters. So, the proposed algorithm was implemented in *R* – a programming language and free software environment for statistical computing and graphics [42].

5.1. EXPERIMENTAL DATA AND SET-UP

In order to be able to compare the performance of our algorithm with those of other algorithms, we had to conduct an experiment using the same conditions as those used in other methods. The conditions involved having the same type of data, the same amount of training data and test data. Experimental data was obtained from a study by Killourhy et al. [43]. The reason for this is, the data is consistent with our formulated static authentication problem, it is representative enough.

We had to also take the experimental set up parameters from the same source. Enrolled for the study were 51 subjects (26 males, 25 females; 35 right-handed and 16 left-handed subjects). They completed eight data-collection sessions (of 50 passwords each), making it 400 password-typing samples in total. We collected password typing data from the 51 subjects who each typed 400 repetitions of a password. We then extracted various timing features, such as dwell time, *press–press* time, *release–press* time, *release–release* time, etc.

Moving further, some assumptions were made. We considered a situation where a user's long-time password has been compromised by an impostor. We assume that the legitimate user is practiced in typing his/her password, while the illegitimate user is not (for example, he is typing it for the first time). So, in this case, we measure how well the detection algorithm is able to differentiate between the genuine user's typing and the impostor's typing.

For a start, we designate one of the 51 subjects as the legitimate user, and the rest as illegitimate users. We train our detector and test its ability to identify the genuine user and impostors. So, the training phase of the algorithm is run on the timing vectors from the first 200 password repetitions typed by the genuine user. The algorithm builds a model of the user's typing behavior. Then, the test phase of the algorithm is run on the timing vectors from the remaining 200 password repetitions typed by the genuine user. The anomaly scores assigned to each timing vector are recorded as legitimate user scores. We then run the test phase of the detection algorithm from the first five password repetitions typed by each of the 50 illegitimate users. The anomaly scores assigned to each timing vector are recorded as illegitimate user scores. In total we have 450 attempts.

The above process is repeated, each time designating one of the other subjects as the legitimate user in turn. After training and testing our algorithm, we have a total of 51 sets of legitimate user and illegitimate user scores.

This experimental model corresponds rather precisely to the real scenario of using user authentication based on keystroke dynamics analysis: the first attempts are used for training, assuming that training occurs at the moment when the password changes to a new one, when a legitimate user is just beginning to develop his characteristic password typing traits. For detection, the last attempts are used, where the legitimate user exhibits the developed password typing traits, while the illegitimate ones, being previously unfamiliar with the password, do not.

5.2. PARAMETER SELECTION AND RESULTS

In order to evaluate the influence of all parameters and select the values that best fit the problem, several series of experiments were conducted where we varied parameter values within given intervals. Results were evaluated and appropriate conclusions on how a particular parameter affected the result were reached. After selecting the best value of one parameter, we fixed it and started selecting the value of the next parameter. The variable parameters are further described below in the order in which they were selected.

5.2.1 KERNEL AND ITS PARAMETERS

In our experiments, we adopted the two most popular functions as kernels: dot product and Gaussian kernel as shown in expressions (1) and (3), respectively.

With dot product being used as the kernel, the best result was obtained at CER = 0.32. This indicates that the algorithm performs very poorly when the dot product is used as a kernel. For the Gaussian kernel, we had to vary its standard deviation, sigma σ. For small σ, there was a slow increase in correctly recognized attempts as the threshold decreased. With a strong increase in σ from the optimal value, there was a general degradation of the ROC curve, without any characteristic features; at further increase, the iteration algorithm stopped converging, which, most likely, was due to rounding errors inherent in calculations involving floating-point numbers. The optimal value of σ for the presented sample turned out to be 101.

5.2.2 DISTANCE FROM THE CLUSTER CENTER

As described earlier, there are two ways (simplified and iterative) to find η, which is the distance from the cluster center at which the degree of membership is assumed to be 0.5. When using the simplified method, no additional parameters are required. The best CER obtained using the simplified method was 0.187. For the iterative method, the outlier proportion must be specified. Experimentally, the best expected outlier proportion was found to be 0.1. Using the iterative algorithm, it gives a CER of 0.177. When the outlier proportion is varied between 0.05 and 0.2, the CER varies between 0.181 and 0.177.

5.2.3 DATA NORMALIZATION FACTORS

The most significant improvements in performance were obtained after normalizing the input data before processing. The reason for this is that the parameters, by their nature, have very different values. For example, the dwell time of a key is always strictly positive, while the interval between a key release and the next key press time may be positive or negative. Therefore, some normalization factors were considered, namely normalization to the square root of the variance, the absolute deviation, the square root of the absolute deviation, the interquartile range, and median absolute deviation. The absolute deviation and its square root gave the most accurate results for normalization.

A point that should be mentioned here is that for some users, the first type of normalization turned out to be better, and for others the second type. In this regard, we attempted to find out the best normalization method at the time of training. Cross validation was used for this purpose. The training sample was divided into two halves – the first part was normalized to the absolute deviation and to the root of the absolute deviation. Then two models were trained on it and tested on the second half of the training sample, respectively. After the results were obtained, training was similarly performed on the second half and testing on the first half. The best normalization method for the user was the one with the smaller mean CER. However, similarly, introduction of cross validation did not improve the CER.

Additionally, when processing data, we replaced each value x in the input data with natural logarithm *ln(x+C)*, where the value of *C* was taken as large as possible, such that *ln(x+C)* could be computed. This decision was justified by the fact that random variables describing individual password typing features, as was found out during the experiment, were more or less lognormally distributed.

6. RESULTS AND DISCUSSION

The chart in figure 2 compares the overall performance of some of the algorithms considered in [43] with the suggested algorithm. The algorithms have been rank-ordered in alphabetical order. In the chart, our proposed algorithm is designated as "SUGGESTED ALGORITHM".

The suggested algorithm obtained the lowest crossover error rate (0.093), thus indicating higher accuracy and reliability. The Manhattan (scaled), Nearest Neighbor (Mahalanobis), and the Outlier Count (z-score) detector were the other top-performing detectors using the crossover error performance measure. Our algorithm turned out to be 0.003 better than the best performer – the Manhattan (scaled), at 0.096.

The anomaly-detection algorithms were appraised based on the same data, under the same conditions, and using the same procedures. Therefore, differences in performance can be credited to the algorithm and not to different experimental conditions.

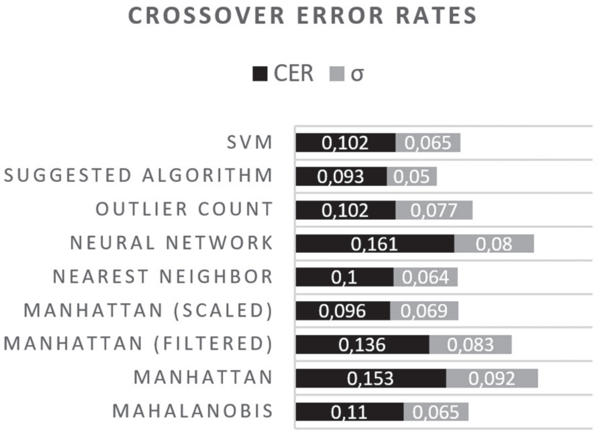


Fig. 2. A comparison of the performances of anomaly-detection algorithms

A critical challenge here is that minor differences in the algorithms and even in the assessment can trigger substantial changes in performance. Typing biometrics is a delicate instrument in a noisy domain. As long as assessment and comparison depend on controlling these small differences in performance, shared data and similar assessment procedures are crucial. So extra shared data and further assessments are required to determine and disentangle the factors facilitating and hindering the performance of each algorithm.

To further validate obtained results, the zero-miss false-alarm rate (ZM-FAR) could be computed and compared across the anomaly-detection algorithms considered. To calculate the ZM-FAR, the threshold is chosen such that FAR is minimized under the constraint that the miss rate be zero. This measure was used in some earlier studies [44, 45]. The CER and ZM-FAR are different performance measures, but both are error rates (i.e., lower values imply fewer errors and better performance).

7. CONCLUSION

In this work, we have investigated keystroke biometrics-based static authentication problem. In doing so, we proposed a new model – based on fuzzy sets and kernel function – for representing raw keystroke data, developed and implemented the corresponding algorithm, and compared it with some classic anomaly-detection algorithms, via experiments, on an equal basis. Our suggested method was found to have outperformed existing methods (with respect to the static authentication problem) – obtaining the lowest crossover error rate.

We have made some trade-offs, which certainly influenced performance. For this reason, the data could be used to assess what impact such decisions have. To give an example, we used 200 samples for training, which may seem unrealistically large. Moreover, we used unpracticed illegitimate users, which appears to be impractical because such users might practice if they knew timing mattered, thereby enhancing detector performance. We have made these trade-offs for the sake of unbiased assessments.

To achieve high performance with less training data, a different appraisal procedure could be adopted to train the detection algorithm using fewer passwords. However, such should be categorically and rigorously described to avoid conflating and confusing different appraisal methods.

8. REFERENCES

[1] W. H. Yang, S. P. Shieh, "Password authentication schemes with smart cards", Computers & Security, Vol. 18, No. 8, 1999, pp. 727–733.

[2] D. S. Carstens, P.R. McCauley-Bell, L.C. Malone, R.F. DeMara, "Evaluation of the human impact of password authentication practices on information security", Information Science Journal, Vol. 7, No. 1, 2004, pp. 67–85.

[3] I. Sluganovic, A. Karlovic, P. Bosilj, M. Šare, S. Horvat, "User authentication based on keystroke dynamics analysis", Proceedings of the 35th International Convention MIPRO, Opatija, Croatia, 21-25 May 2012, pp. 2136–2141.

[4] M. Styugin, "Dynamic key password authentication", International Journal of Security and Networks, Vol. 14, No. 2, 2019, pp. 78-85.

[5] G. Shafi, S. Micali, and R. L. Rivest, "A digital signature scheme secure against adaptive chosen-message attacks", SIAM Journal on Computing, Vol. 17, No. 2, 1988, pp. 281–308.

[6] U. Maurer, "Intrinsic limitations of digital signatures and how to cope with them", Proceedings of the 6th International Conference on Information Security, Bristol, UK, 1-3 October 2003, pp. 180–192.

[7] A. Levi, C. B. Güder, "Understanding the limitations of S/MIME digital signatures for e-mails: A GUI based approach", Computers and Security, Vol. 28, No. 3-4, 2009, pp. 105–120.

[8] RecFaces, “What is biometric security and why does it matter today?”, https://recfaces.com/articles/biometric-security (accessed: 2021).

[9] M. Chrobok, Physical Biometrics vs Behavioral Biometrics, https://www.revelock.com/en/blog/physical-biometrics-vs-behavioral-biometrics (accessed: 2021).

[10] I. Alsaadi, "Study on most popular behavioral biometrics, advantages, disadvantages and recent applications: a review", International Journal of Scientific & Technology Research, Vol. 10, No. 1, 2021, pp. 15–21.

[11] M. Bača, M. Schatten, J. Ševa, "Behavioral and physical biometric characteristics modeling used for ITS security improvement", Transport Problems, Vol. 4, No. 4, 2009, pp. 5–13.

[12] T. Grance, M. Stevens, M. Myers, "Guide to Selecting Information Technology Security Products", NIST Guide to Selecting Information Technology Security Products (NIST Special Publication 800-36), 2003.

[13] P. S. Teh, A. B. Teoh, S. Yue, "A survey of keystroke dynamics biometrics", The Scientific World Journal, Vol. 2013, p. 408280, 2013.

[14] M. S. Obaidat, "Verification methodology for computer systems users", Proceedings of the 1995 ACM Symposium on Applied Computing, Nashville, Tennessee, USA, 26-28 February 1995, pp. 258–262.

[15] V. N. Gudivada, V. V. Raghavan, V. Govindaraju, C. R. Rao, "Handbook of Statistics", Cognitive Computing: Theory and Applications, Vol. 35, 2016, pp. 2–384.

[16] H. F. Tipton, M. Krause, "Information Security Management Handbook", Auerbach Publications, 6th Edition CRC Press LLC, 2007.

[17] E. Conrad, S. Misenar, J. Feldman, "CISSP Study Guide, Syngress", 2nd Edition, 2012, p. 599.

[18] E. Conrad, S. Misenar, J. Feldman, "Eleventh Hour CISSP: Study Guide", Syngress, 3rd Edition, 2017, p. 200.

[19] F. W. M. H. Wong, A. S. M. Supian, A. F. Ismail, "Enhanced user authentication through typing biometrics with artificial neural networks and K-nearest neighbor algorithm", Proceedings of the 35th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 4-7 November 2001, pp. 911–915.

[20] E. Lau, X. Liu, C. Xiao, X Yu, "Enhanced user authentication through keystroke biometrics", Computer and Network Security, Vol. 6, 2004.

[21] F. Bergadano, D. Gunetti, C. Picardi, "User Authentication through keystroke dynamics", ACM Transactions on Information and System Security, Vol. 5, No. 4, 2002, pp. 367–397.

[22] G. Saggio, G. Costantini, M. Todisco, "Cumulative and ratio time evaluations in keystroke dynamics to improve the password security mechanism", Journal of Computer and Information Technology, Vol. 1, No. 2, 2011, pp. 4–11.

[23] R. Solanki, P. Shukla, "Estimation of the user's emotional state by keystroke dynamics", International Journal of Computer Applications, Vol. 94, No. 13, 2014, pp. 21–23.

[24] K. S. Sung, S. Cho, "GA SVM wrapper ensemble for keystroke dynamics authentication", Proceedings of the International Conference on Biometrics, Hong Kong, China, 5-7 January 2006, pp. 654-660.

[25] M. I. Petrovsky, "Outlier detection algorithms in data mining systems", Programming and Computer Software, Vol. 29, No. 4, 2003, pp. 228–237.

[26] B. Schölkopf, A. J. Smola, "A short introduction to learning with kernels", Advanced Lectures on Machine Learning, Lecture Notes in Computer Science, Springer, 2003.

[27] T. Hofmann, B. Schölkopf, A. J. Smola, "Kernel methods in machine learning", The Annals of Statistics, Vol. 36, No. 3, 2008, pp. 1171–1220.

[28] Y. Cho, L.K. Saul, "Kernel methods for deep learning", Proceedings of the 22nd International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 7-10 December 2009, pp. 342-350.

[29] H. Chiroma, S. Abdulkareem, A.I. Abubakar, T. Herawan, "Kernel functions for the support vector machine: comparing performances on crude oil price data", Recent Advances on Soft Computing and Data Mining, Springer, 2014, pp. 273–281.

[30] C. Savas, F. Dovis, "The impact of different kernel functions on the performance of scintillation detection based on support vector machines", Sensors, Vol. 19, No. 23, 2019, p. 5219.

[31] M. A. M. Hasan, S. Xu, M. M. J. Kabir, S. Ahmad, "Performance evaluation of different kernels for support vector machine used in intrusion detection system", International Journal of Computer Networks and Communications, Vol. 8, No. 6, 2016, pp. 39–45.

[32] F. Meng, Y. Fu, F. Lou, Z. Chen, "An Effective Network Attack Detection Method Based on Kernel PCA and LSTM-RNN", Proceedings of the International Conference on Computer Systems, Electronics and Control, Dalian, China, 25-27 December 2017, pp. 568–572.

[33] Z. Rustam, N. Olievra, "Comparison of fuzzy robust Kernel C-Means and support vector machines for intrusion detection systems using modified kernel nearest neighbor feature selection", Proceedings of the 3rd International Symposium on Current Progress in Mathematics and Sciences, Bali, Indonesia, 26–27 July 2017.

[34] N. Baili, H. Frigui, "Fuzzy clustering with multiple kernels in feature space", Proceedings of the IEEE International Conference on Fuzzy Systems, Brisbane, QLD, Australia, 10-15 June 2012, pp. 1–8.

[35] L. A. Zadeh, "Fuzzy sets", Information and control, Vol. 8, No. 3, 1965, pp. 338–353.

[36] D. Klaua, “About an approach to multi-valued set theory”, Monatsblatt Deutscher Akademie der Wissenschaft, Berlin, 1965, pp. 859–876. (in German)

[37] C. Kahraman, B. Öztayşi, S. Çevik Onar, "A comprehensive literature review of 50 years of fuzzy set theory", International Journal of Computational Intelligence Systems, Vol. 9, Sup 1, 2016, pp. 3–24.

[38] Y. Nesterov, Efficiency of coordinate descent methods on huge-scale optimization problems, SIAM Journal on Optimization, Vol. 22, No. 2, 2010, pp. 341–362.

[39] P. Richtárik, M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function", Mathematical Programming, Series A, Vol. 144, No. 1-2, pp. 1–38.

[40] A. Ene, H. L. Nguyen, “Random coordinate descent methods for minimizing decomposable submodular functions”, Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6-11 July 2015, pp. 787–795.

[41] R. Wang, C. Chen, J. Lee, E. F. Darve. "PBBFMM3D: A parallel black-box algorithm for kernel matrix-vector multiplication", Journal of Parallel and Distributed Computing, Vol. 154, 2019, pp. 64–73

[42] R Core Team. “R: A language and environment for statistical computing. R Foundation for Statistical Computing”, Vienna, Austria. http://www.R-project.org/ (accessed: 2020)

[43] K. S. Killourhy, R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics", Proceedings of the International Conference on Dependable Systems & Networks, Estoril, Lisbon, Portugal, 29 June - 2 July 2009, pp. 125-134.

[44] S. Cho, C. Han, D. H. Han, H. Kim, "Web-based keystroke dynamics identity verification using neural network", Journal of Organizational Computing and Electronic Commerce", Vol. 10. No. 4, 2000, pp. 295–307.

[45] E. Yu, S. Cho, "GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification", Proceedings of the International Joint Conference on Neural Networks, Portland, OR, USA, 20-24 July 2003, pp. 2253–2257.

A Novel Method for Real Time Protection of DC Microgrid Using Cumulative Summation and Wavelet Transform

Original Scientific Paper

Dipti D Patil
Department of Electrical Engineering, Fr. C. Rodrigues Institute of Technology
Navi Mumbai, Maharashtra, India
diptipatil009@gmail.com

Bindu S
Department of Electrical Engineering, Fr. C. Rodrigues Institute of Technology
Navi Mumbai, Maharashtra, India
bindu.s@fcrit.ac.in

Sushil Thale
Department of Electrical Engineering, Fr. C. Rodrigues Institute of Technology
Navi Mumbai, Maharashtra, India
sushil.thale@fcrit.ac.in

Abstract *– DC microgrid is a compact framework comprising interconnected nearby sources and loads. The renewable energy source used in DC microgrids being intermittent leads to the change in the power availability as well as the fault current levels. In such situations, detecting and clearing the faults is very important to protect the DC microgrid without compromising on fault clearing time and interruption of the load. This paper proposes a hybrid Cumulative Sum (CumSum) and Wavelet transform-based approach to detect the fault. The CumSum value raises the amplitude by averaging the fault current. Wavelet transforms obtain important fault current features by decomposing the current signal. The hybrid method of CumSum and Wavelet analysis proposed here enables the detection of the fault and differentiates the fault condition from sudden load variation. Additionally, it helps to recognize the location of the fault by the wavelet energy difference. The proposed scheme is tested with a developed ring-type low voltage DC (LVDC) microgrid hardware model under various fault conditions. The scheme is implemented using TMS320F28069 digital signal processors (DSP) of Texas Instruments. The hardware results are validated using MATLAB simulation. The proposed method performance is also compared with the existing methods used for DC microgrid protection. The outcome shows that the proposed method has a high accuracy of 98.72%, selectivity of 96.08%, and reliability of 99.01%. The execution time required by the proposed method is also less.*

Keywords*: Fault Detection and protection, DC microgrid, CumSum, Wavelet Transform.*

1. INTRODUCTION

The microgrid (MG) is a rising and most encouraging idea in both AC and DC systems. It can be mainly classified into AC, DC, or a mix of both i.e. hybrid. The microgrid comprises the small interconnection of local distributed renewable and non-renewable sources and loads. Such a structure is efficient for working in one islanded and the other grid-associated modes [1]. A microgrid provides a reliable and high-power quality electric supply as well [2]. In a DC microgrid, a large portion of the sustainable power sources are DC sources like solar, fuel cell, and so on that can supply to the DC load directly. Hence eliminating the conversion stages and improving the overall system efficiency. Control requirements are also reduced in DC microgrids because of the absence of reactive power control and frequency synchronization as compared to the AC microgrid [3]. However, along with these benefits, microgrids have also raised several challenges like high installation cost and protection issues, and low system inertia [4]. Because of the discontinuous type of the sources, low fault levels, and switching structure arrangement, protection becomes one of the significant difficulties in a microgrid. This poses a constraint on the fault detection time and isolation of the healthy part. Also, in the AC system, fault location is identified from the variation of impedance. In the case of DC structure, fault impedance is not the same as that of the AC structure because of the smaller length of the lines and thus identification of the exact location of the fault is additionally a difficult issue [5]. In DC microgrid sources and load are connected through power electronic converters. These converters are operated in voltage control mode or current control mode to maintain the output power constant. The output of the power converter on the load side is constant. Therefore a reduction in voltage of the power converter leads to an increase in load current which results in oscillations in voltage and current [6]. In the case of standalone operation of DC microgrid, the fault current is low, it will become difficult to protection devises to distinguish between oscillation due to fault scenario or oscillation due to constant power operation of the converter.

Adaptive protection of the distribution system is proposed by Mahat, *et al*. using directional overcurrent relays. Relays have fixed commanding attributes for grid-associated and island conditions and operate only for the forward direction of the fault current [7]. A power probe unit equipped with a non-iterative strategy was proposed to find fault position for LVDC microgrid. However, this technique utilizes additional instruments to estimate the DC fault area [8]. Active impedance and a traveling wave-based fault identification technique are the few methods used for DC microgrid for protection [9-10]. Balasreedharan, *et al*., and Meghawani, *et al*. proposed higher-order derivatives for the detection of the fault but this method is immensely delicate to the noise amount of the signal [11-12]. The current differential protection (CDP) proposed in the literature requires communication devices that increase the weight size and cost of the system. Anyways, the ongoing improvement of the smart grid extends the provision for the incorporation of a more particular, communication-based, protection strategy that uses this high-level infrastructure. Also, the output of the current transducer used for DC measurement is the voltage which allows an effortless combination with digital processing devices [13]. Recently, most of the research that relied on time frequency analysis to find the fault features and these features are as attributes to learn and build classifier model. This classifier model was then used for detection and classification of faults. However, the performance of the classifier model, which was based on time frequency analysis, was significantly affected by the computational complexity and time of technique [14]. Deep neural network enabled fault detection in LVDC microgrid using empirical mode decomposition (EMD) is proposed by Dipti et al. In this paper EMD extracts the characteristics of the fault signal and these extracted features are utilized at the time of training of the Convolutional neural network (CNN). The trained network is then utilised to classify the normal, abnormal and fault condition in LVDC microgrid. [15].

DC microgrid fault current level undergoes enormous alteration because of the uncertain characteristics of the intermittent sources, changes in network topology, and switching on and off of heavy or light load. Therefore, traditional overcurrent protection schemes do not work effectively against the fault in the DC microgrid. Taking the above issues into account, the contribution of the proposed exploration method related to DC microgrid protection is differentiating between system dynamics and fault scenario, the quick discovery of the fault, and the disconnection of the defective segment to protect healthy sections from high fault current surge. As well as a similar procedure can be utilized to recognize the specific area of a fault. A real-time fault identification technique using Cumulative Sum and wavelet-based characteristics extraction of fault current wave in a DC microgrid is proposed in this paper. This paper likewise features the fault location identification from the Energy per cycle of the wavelet coefficient.

The paper is coordinated as follows; an audit of existing protection techniques for the microgrid is accounted for in section 1. Possible fault in the DC microgrid and fault current behavior is presented in section 2. Wavelet-based fault detection and localization using fault current features and calculation of the cumulative sum of fault current are explained in section 3. The Proposed fault detection algorithm using *CumSum* average, wavelet transform analysis and threshold setting is presented in the section. 4. The development of a lab model of an LVDC microgrid is presented in section 5. The suggested protection scheme is tested on hardware and results are discussed in section 6. Hardware results are approved using simulation results under different fault situations are presented in section 7 followed by the conclusion in section 8.

2. DC MICROGRID FAULT ANALYSIS

DC Bus to bus and bus to ground faults mainly occur in the DC microgrid system. When the positive and negative bus comes in contact, it creates a direct short-circuit fault or bus to bus fault and when the positive or negative bus comes in contact with the ground, it develops the bus to ground fault. In a DC microgrid bus to bus fault is the major serious fault for the converter. As a power electronic converter can be obstructed for self-security during faults; it allows reverse diodes of the device to be exposed to a fault [16]. A bus-to-bus fault in the DC segment may be responsible for peak transient fault current because of charged capacitors and the short impedance of the segment. Figure 1 is the equivalent circuit representation of the DC segment under fault conditions. Voltage source converter (Vsc) is the voltage source converter connected across the source. The notations r and L are used to represent the resistance and inductance of the segment. Filter capacitance is represented by notation C and Rf as well as If are the fault resistance and fault current respectively. During the bus to bus fault, the capacitor starts discharging through the segment impedance of the source to bus fault the point(R = r + Rf and L).

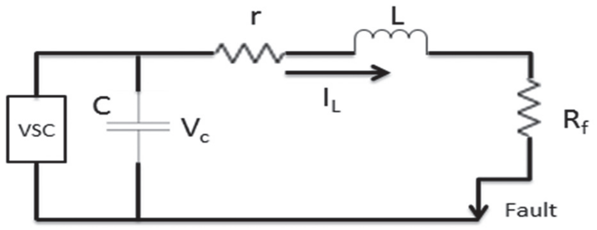


Fig. 1. RLC Equivalent fault circuit

$$i(s) = \frac{\frac{Vc(0)}{L} + i_L(0)s}{s^2 + \frac{R}{L}S + \frac{1}{LC}} \tag{1}$$

Where, $V_C(0)$ is the voltage across the capacitor and $i_L(0)$is the current through the inductor just before the occurrence of fault. The equation of fault current in time domain is

$$i(t) = \frac{V_C(0)}{L\,(S_2 - S_1)}[e^{-S_1 t} - e^{-S_2 t}] + \frac{i_L(0)}{S_2 - S_1} \tag{2}$$

$$[-S_1 e^{-S_1 t} + S_2 e^{-S_2 t}]$$

where

$$S_1, S_2 = \frac{\mathrm{R}}{2\mathrm{L}} \pm \sqrt{(\tfrac{R}{2L})^2 - (\tfrac{1}{LC})} \tag{3}$$

$$S_1, S_2 = \alpha \pm \sqrt{(\alpha)^2 - (\omega o)} \tag{4}$$

Where α is the attenuation and ω_0 is the damped resonant frequency of the fault current. S_1 and S_2 represents the roots of the equation 3. Depending on the value of $(R/2L)^2$, $(1/LC)$ the values of S_1 and S_2 become real or complex and the fault current response can be over damped, critically damped and under damped. Equation 4 uses notation Hence the rate of oscillation of the fault current relies upon segment parameters R, L, and filter capacitance C [18].

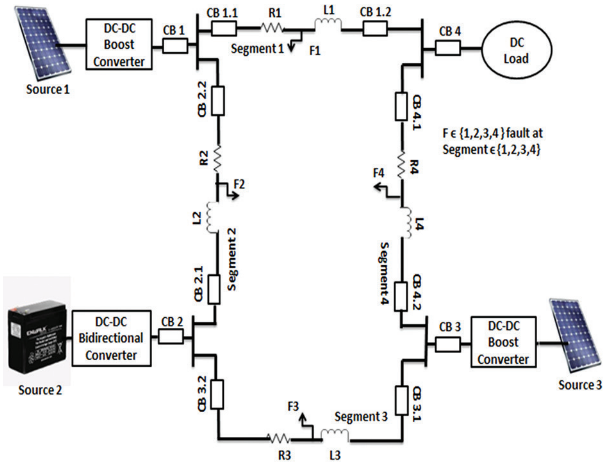


Fig. 2. Simulation model for 48 V DC Microgrid

The 48-V dc ring-type architecture shown in Figure 2 with specifications is given in Table II. The DC microgrid model consists of four segments and the length of each segment is taken into an extent of a kilometre. The system is consisting of four nodes; these nodes are connected with two Renewable Energy Sources (RES), a lead-acid battery bank, and DC electronic load respectively. Each source is interfaced to the microgrid via power electronic converters. (Parameter of converters is same as shown in Table II) Source 1 and source 3 are Solar Photovoltaic (SPV), sources that are interfaced to DC microgrid through a boost converter. The control loop operation of the converter is implemented through the maximum power point tracking (MPPT) controller shown in figure 3. Source 2 is an energy storage system (ESS) source that is interfaced to the DC microgrid through a bi-directional converter. The state of charge (SOC) control loop is executed internally as shown in figure 4. To detect and isolate DC microgrid against fault, relays and circuit breakers are placed at each end of all segments.

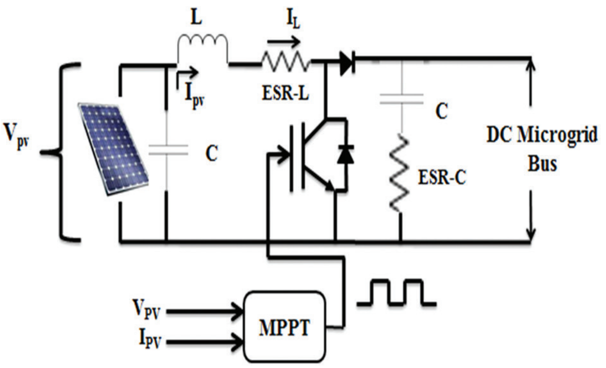


Fig. 3. Control schematic of SPV Source 1 and 3 fed DC microgrid via boost converter

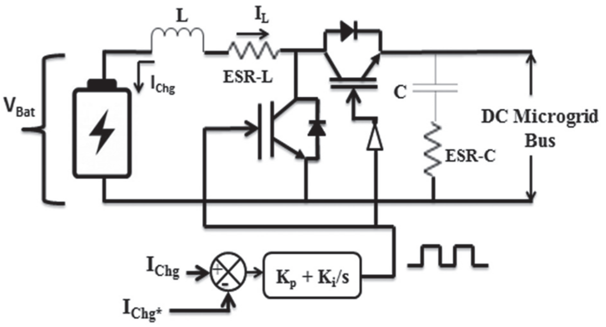


Fig. 4. Control schematic of Battery Source 2 fed DC microgrid via bidirectional converter

The fault is created in segment 1 at 10% and 90% distance of the 1-km bus portion with respective source 1. The fault current has a fast-rising transient with a large peak followed by oscillation. The amplitude and oscillation frequency of fault current is different at both distances as it depends on the equivalent parameters of the fault path seen in Figure 5. The distinction of fault current is estimated in the segment by embedding diverse fault resistance (0.1Ω and 0.9Ω) at the half distance of a 1-km bus segment as displayed in Figure 6. The damping effect is marginally high in the case of 0.9 Ω fault resistance compared to 0.1 Ω resistances.

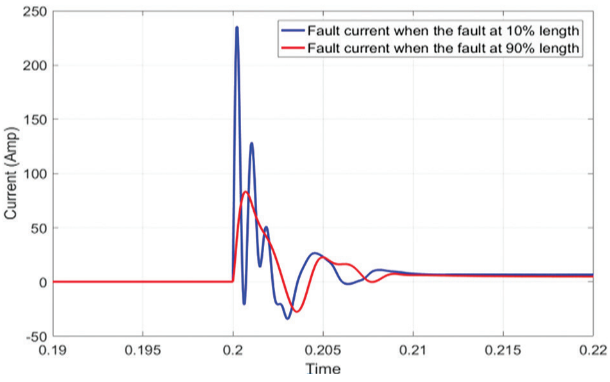


Fig. 5. Difference of rate of rise of fault current after the fault is initiated at 10% and and 90% distant in segment

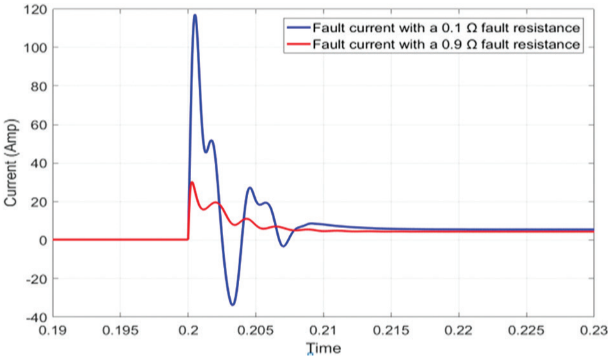


Fig. 6. Difference of rate of rise of fault current after the fault is initiated at half distance of a 1-km bus segment with 0.1 Ω and 0.9Ω fault resistance

3. PROPOSED FAULT DETECTION AND LOCALISATION METHOD

A hybrid Wavelet and cumulative sum based algorithm is proposed to detect transient and moving average of fault current.

3.1 CUMULATIVE SUMMATION

This technique is a mix of the sample by sample comparison method and moving average cumulative summation (*CumSum*) approach. During the fault condition the segment current changes essentially. For a successful fault detection technique, certain changes should get recognized online with the least noticing samples or time. *CumSum* is based on a moving average calculated by considering 'n' numbers of samples in a window. The average value of summation of n number of samples significantly decreases the computational burden as it uses the simple summation and subtraction operators [19].

CumSum = Sum total around the window

$$CumSum = \sum_{k=(t-w)}^{t} i(k) \tag{5}$$

Where t is the present time, w is the window dimension for moving average and k is a dummy index used for average estimation. The time delay added into the system is W times the sampling period [20]. The window size is kept the same as the computational time for the wavelet analysis also. The *CumSum* is a moving average of n number of samples hence, does not affect due to underdamped or overdamped response of fault current. *CumSum* raises the amplitude by averaging the fault current so that the protection relay can detect and isolate the faulty segment [21].

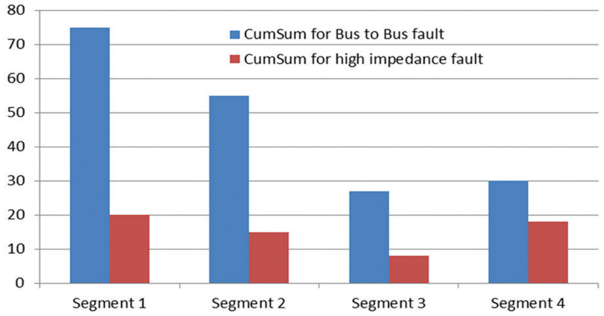


Fig. 7. *CumSum* of fault currents in all segments under bus to bus and high impedance fault in segment 1 of Fig. 2.

Figure 7 shows the variation of *CumSum* of fault current in all segments under the bus to bus and high impedance fault (inserted 1 Ω resistance in fault path) at the end of segment 1 of DC microgrid shown in Figure 2. But, the fault in one segment may increase the *CumSum* of current in neighbouring segments, and it can falsely detect as a fault in the neighbouring segment also (mostly observed in high impedance faults). Also a delay in fault detection is observed when more samples are taken for calculating the average.

3.2 WAVELET TRANSFORM

Discrete Wavelet Transform (DWT) calculation is utilized for computerized execution of constant wavelet change utilizing a two-channel perfect reconstruction filter bank. The information signal is disintegrated into low and high-frequency. The high-pass channel relies upon the mother wavelet work, it estimates detail coefficient ($D1$), and the low pass channel from the scaling capacity of the mother wavelet work estimates approximate coefficients ($A1$) at the main level [22].

$$\mathrm{A_1(n) = \sum_k h(k-2n)x(k)} \tag{6}$$

$$D_1(n) = \sum_k g(k-2n)x(k) \tag{7}$$

Here *h(n)* is high-pass and *g(n)* is a low pass filter coefficient. The input signal is broken down into detail and approximate coefficients through low pass and high pass filters [23]. The subsequent signal is the portrayal of a similar signal in the distinctive frequency groups. Figure 8 shows the wavelet coefficient of fault current in all segments under the bus to bus and high impedance fault (inserted 1 Ω resistance in fault path) at the end of segment 1 of the DC microgrid shown in Figure 2. It is very clear from the bar diagram that the wavelet coefficients of only the faulty segment are much higher than other segment coefficients.

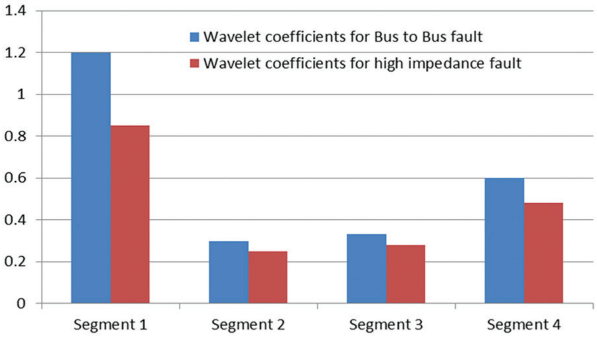


Fig. 8. Wavelet coeffs. of fault currents in all segments under bus to bus and high impedance fault in segment 1(Fig. 2)

However, the dynamics due to a sudden change in load or change in network configuration wavelet analysis gives the magnitude of the same coefficients as that of the magnitude during a fault in the segment. During normal conditions sudden change in load is applied on segment 4 (Figure 2), the load is increased suddenly by 10% of connected load at 0.3 sec. This results in a change in *CumSum* and wavelet coefficient of segment current, it is observed that the wavelet coefficient magnitude is the same as that of the magnitude during a fault in the segment. However, *CumSum* shows only a small variation as shown in Figure 9.

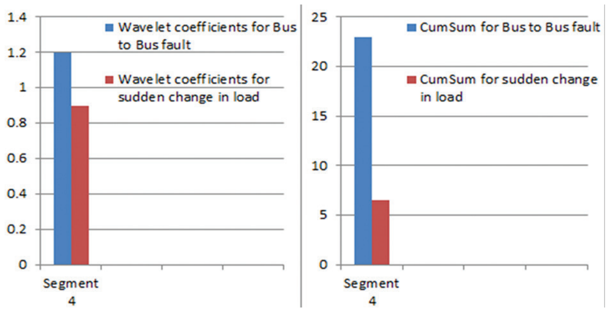


Fig. 9. Wavelet coefficients and *CumSum* of current for Bus to Bus fault and after sudden application of load at Segment 4 (Fig 2)

3.3 FAULT LOCALISATION

The fault location can be recognized from the current measured from the faulty segment. The sudden variations in the DC system can be analysed as RLC transient conditions due to the resistive, inductive, and capacitive parameters of the DC cable. Wavelet transform decomposes this transient signal with multiple resolutions and gives information about fault current energy concentration in a different band. The energy of the wavelet coefficients usually needs normalization [24].

$$Normalisedd_{ij} = \frac{d_{ij}}{A_{ij}} \qquad (8)$$

The energy per cycle from wavelet coefficients [25] is given by

$$E_i = \sum_{j=1}^{n} Normalised\,|d_{ij}|^2 \qquad (9)$$

Where, i = Level of the detail coefficient and j = Samples utilized for each detail coefficient

The detail coefficient extent changes due to the system transients. Consequently; the adjustment of energy demonstrates an unsettling influence in the system. Change in energy is the distinction in energy between two back-to-back cycles it is calculated [26] by equation 10.

$$Change\ in\ Energy = \sqrt{(E_A - E_B)^2} \qquad (10)$$

Where, E_B and E_A represent the energy of the current cycle and the energy of the past cycle respectively. As the fault location changes the energy concentration value between the start and the endpoint of the segment also changes. This change in energy is used to find the fault location in the segment. Fault near the source causes a high amplitude transient which results in a high wavelet coefficient and higher value of energy. If the fault occurs at the end of the segment amplitude of the transients is low and gives less wavelet high pass coefficient and lesser change in energy value.

4. PROPOSED FAULT DETECTION ALGORITHM

The proposed detection method uses two parameters; *CumSum* and wavelet transform. DC Current in the segment is continuously sampled and monitored by the respective current transformer (CTs). *CumSum* and wavelet transform coefficients are calculated from the measured current shown in Figure 10.

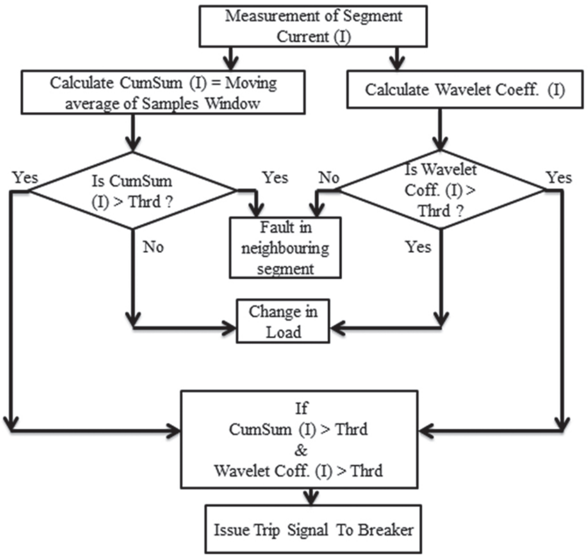

Measurement of Segment Current (I)
Calculate CumSum (I) = Moving average of Samples Window
Calculate Wavelet Coeff. (I)
Yes
Is CumSum (I) > Thrd ?
Yes
No
Is Wavelet Coff. (I) > Thrd ?
Yes
No
Fault in neighbouring segment
Yes
Change in Load
If CumSum (I) > Thrd & Wavelet Coff. (I) > Thrd
Issue Trip Signal To Breaker

Fig. 10. Flow diagram of hybrid protection Scheme

The calculated *CumSum* and wavelet coefficients are compared with the pre-characterized limits of the *CumSum* and wavelet coefficient. The condition, at which the determined *CumSum* and wavelet coefficient are higher than the individual set limit, is distinguished as the fault and a trip signal is initiated to separate the inoperative section from the leftover robust system. If only the *CumSum* of segment current becomes greater than the threshold, it shows a fault in the neighboring segment and if only calculated wavelet coefficients of segment current become greater than threshold indicates a sudden change in load or network configuration.

Table 1. Threshold (*CumSum* and wavelet coefficient) settings for all relays

X	*CumSum* (Amp) measured at CB(X.1)	Wavelet Coeffs. Measured at CB(X.1)	*CumSum* (Amp) measured at CB(X.2)	Wavelet Coeffs. Measured at CB(X.2)
1	22	1.4	20	0.9
2	24	0.8	34	6.5
3	20	0.8	34	6.5
4	18	0.9	23	0.8

The threshold of the *CumSum* and wavelet analysis method is determined by initiating fault at the end of the segment with 1Ω fault resistance one after another. The threshold values obtained by this method are tabulated in Table 1. X is the segment number and it varies from one to four. X.1 indicates the threshold corresponding to one end of the segment and X.2 the other.

5. PROTOTYPE OF LVDC MICROGRID

The model of the 48 volts LVDC microgrid is planned and the protection scheme is tested in the hardware setup shown in figure 11.

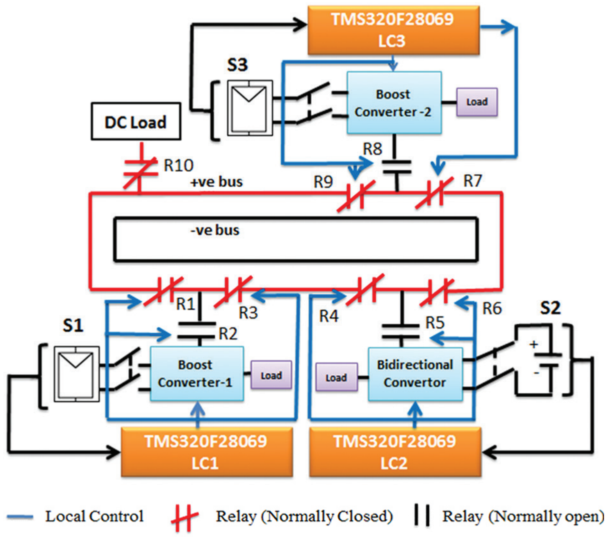


Fig. 11. Designed model of LVDC Microgrid

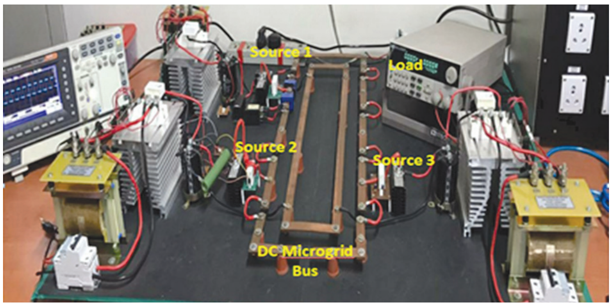


Fig. 12. Photograph of laboratory prototype of the DC microgrid

Table 2. Hardware and simulation parameters used to design dc microgrid

Parameters	Specifications
DC grid voltage	48V
Segment Resistance	10 mΩ / Km
Segment Inductance	100 μH / Km
Segment Length	1 Km
Boost and Bidirectional converter Parameter	Power - 1000W
	Operating frequency - 20KHz
	Switch and Diode -100A, 600V IGBT module

The system is consists of two Renewable Energy Sources (RES), a lead-acid battery bank, and a DC load. Each component of hardware is designed and developed in the same way as the simulation model explained in figure 2. The parameters of the designed converters are shown in Table II. Texas Instruments' TMS320F28069 digital signal processor (DSP) processor is utilized to execute control for converters and switches. Resistive attenuators are used for voltage sensing from the bus nodes. The current is sensed by the hall-effect current sensors (LEM, CTG-FBC, 50A, and 100A). The noise from the measured signal is filtered by passing it through an RC Low Pass filter and fed to the ADC to the sampling of the Digital Signal Processors (DSP's) [27-28]. Source 2 works in voltage control mode to maintain the DC bus voltage, and other sources work in current control mode to share the load current. The DSP of respective sources acts as a local controller for each source. It takes feedback from the sensing circuit and the voltage or current control mode is accomplished. Power switches (R1-R9) are integrated in the microgrid structure to isolate faulty parts under the fault condition. Power electronic load is used as DC load and is connected through a power switch R10 as shown in figure 11. Each converter connected with an appropriate power resistor working as a local load. The experiment setup is shown in figure 12.

6. HARDWARE RESULTS

To assess the performance of hardware it is operated under different circumstances. The suggested protection technique is implemented and verified for the bus to bus and high impedance short circuit faults. All the three sources and loading conditions decide the fault current level of the system. The major challenge in Low voltage DC microgrid is the low fault current level under no grid connectivity condition.

6.1 PROTECTION SCHEME OPERATION UNDER DIRECT SHORT CIRCUIT FAULT

To check the proposed algorithm, a bus to bus fault is initialized at segment 1 of the microgrid at 700 meters from the source. All the sources operated with limited capacity to ensure the safety of switches. As the fault occurred at segment 1, a drop in converter output voltage, a drop in DC bus voltage and rise in load current is observed as shown in Figure 13. The faulty segment is separated by power switches within 150μs and from the calculated energy, it is identified that fault is occurs at 700 meters from the source in the segment.

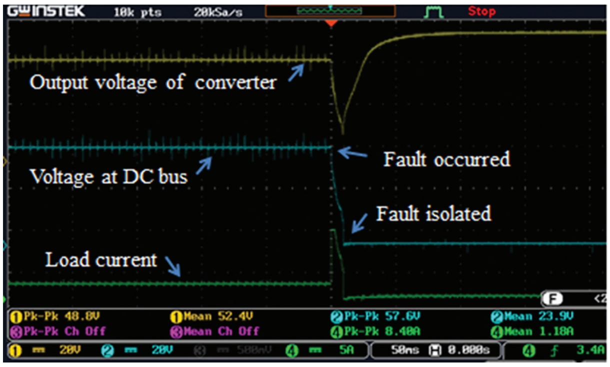
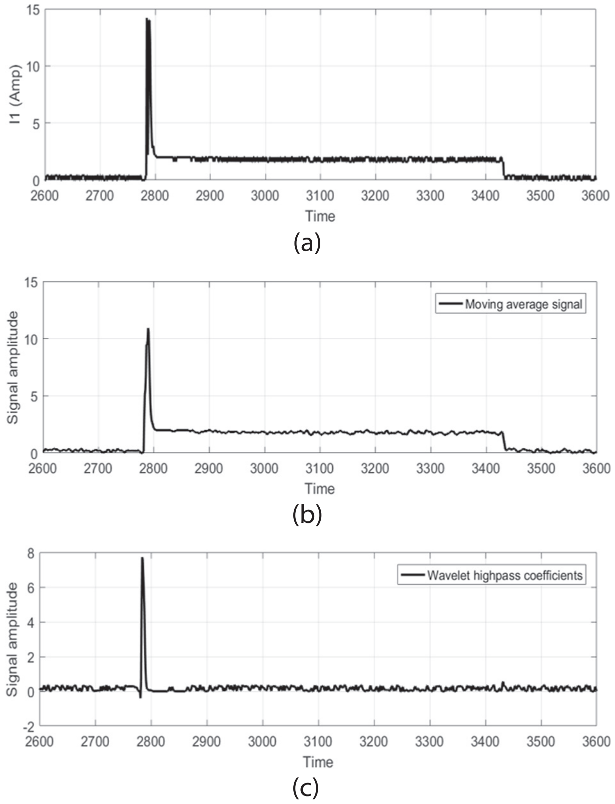


Fig. 13. Converter output voltage, DC bus voltage and Load current when fault occurred in segment 1.

To observe the *CumSum* and wavelet response of hardware result under the bus to bus and high impedance faults, the fault signal is stored into a CSV file from a digital storage oscilloscope and is fed to MATLAB code. Figure 14 shows the results of fault current, its calculated *CumSum*, and wavelet coefficient of fault current under bus to bus fault of segment1. Fault current has high absolute values of 14 amp with a very high rate of change in 0.4 msec.



(a)

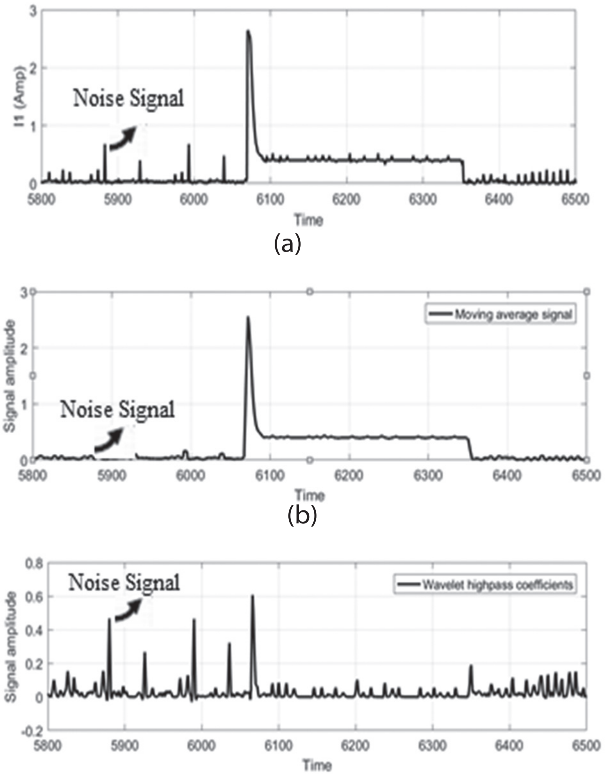
(b)

(c)

Fig. 14. (a) Bus to bus fault current at segment 1 (b) *CumSum* of fault current (c) Wavelet coefficients of fault current.

6.2 PROTECTION SCHEME OPERATION UNDER HIGH IMPEDANCE FAULT

The scheme is checked for the high impedance fault by inserting fault resistance 1Ω in the fault path and keeping other working conditions and fault location same as previous case. The insertion of fault resistance restricts the peak value to 2.8 amp and it takes 0.4 msec. Figure 15 (a) shows the segment 1 fault current, this signal has noise due to the effect of measuring devices and the communication channels used in the system. Cumulative summation of signal in a moving average window acts like a low pass filter which does not affect the slow rise noise in the fault signal (figure 15 (b)). Also, the wavelet transform used high frequency transient to detect the fault (figure 15(c)). Hence the suggested protection technique likewise works even in a noisy environment.



(a)

(b)

(c)

Fig. 15. (a) High impedance fault current at segment 1 (b) *CumSum* of fault current (c) Wavelet coefficient of fault current.

7. SIMULATION RESULTS

The hardware shown in figure 11 is modelled in MATLAB (Figure 2) to verify the feasibility of the algorithm under different fault conditions. The simulation parameters used for modelling are given in Table II.

The system is first simulated without fault condition by considering 1000 W load at nominal load voltage 100 volt. Both the SPV Sources S1, and S3 are contributing 81.42 W and 49.80 W is contributed by battery Source 2. Under normal operation, the source S1 and S3 are supplying the major load. When the load on the microgrid is slightly increased, source S2 start sharing the increased load on the microgrid, bus voltage dips slightly for the period of overload. Figure 16 shows all four segment currents under normal operating conditions.

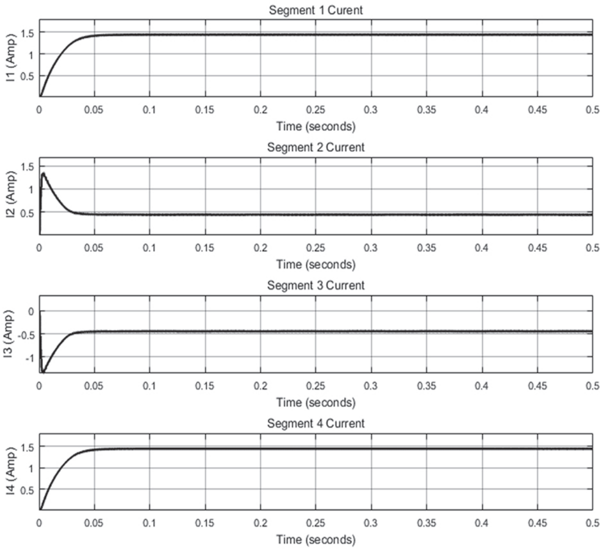
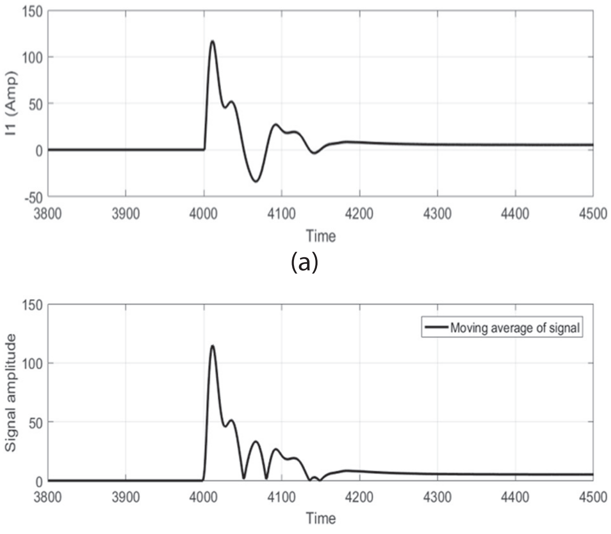


Fig. 16. Segment currents under normal condition

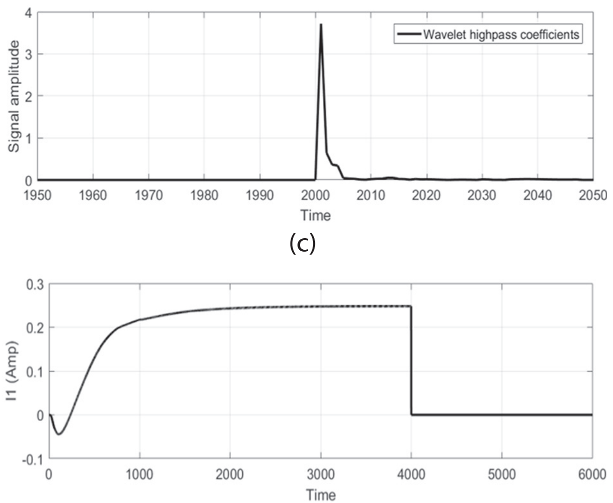
7.1 PROTECTION SCHEME OPERATION UNDER DIRECT SHORT CIRCUIT FAULT

The model is tested for the bus to bus fault in segment 1 at 500 meters shown in figure 17 at 0.2 sec without source limitation. The current raises 120A within 0.5 msec. Due to bus to bus short circuit, high currents transient appear in all the segments (Figure 2). Within a few milliseconds of the fault occurrence, the bus voltage decreases to zero. The segment currents are continuously measured and compared with the threshold. As the fault current exceeds the pre-set value of the calculated wavelet coefficients and the *CumSum* the protection algorithm generates a trip signal to detach the line, as shown in figure 17 (d).



(a)

(b)



(c)

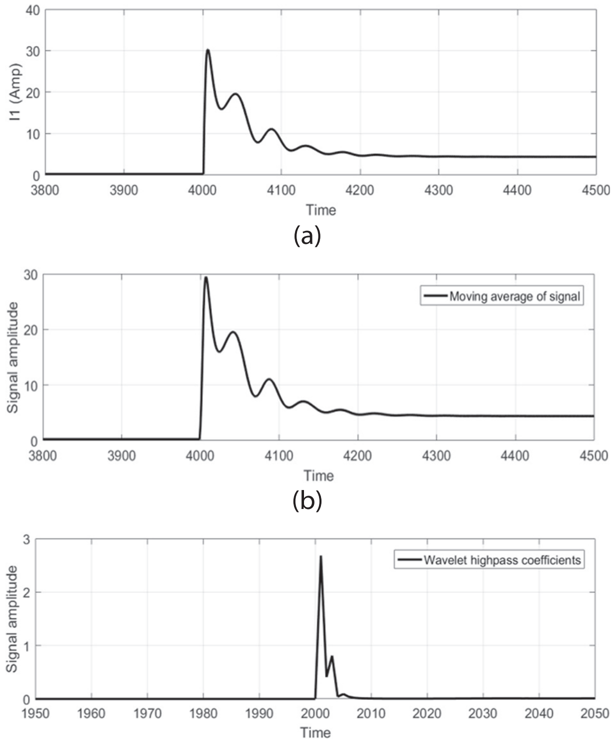
(d)

Fig. 17. (a) Segment 1 current due to bus to bus Fault (b) *CumSum* of fault current of segment 1 (c) Wavelet Coefficients of fault current of segment 1 (d) Segment 1 current after detection of fault.

The direct bus to bus fault is detected within 150µs, which guarantees the life of converters and other framework parts are ensured. The DC bus voltage immediately drops to zero on account of the excessive rate of rise of fault current.

7.2 PROTECTION SCHEME OPERATION UNDER HIGH IMPEDANCE FAULT

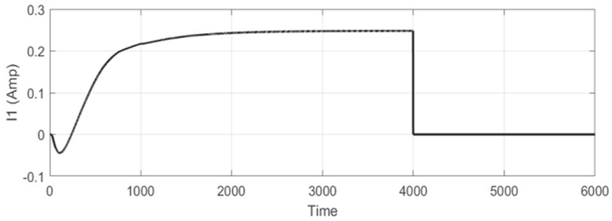
The system is tested for high impedance faults by inserting resistance RF = 1Ω in the fault path by keeping operating conditions and location of the fault stay the same as in the past case.



(a)

(b)

(c)



(d)

Fig. 18. (a) Segment 1 current due to high impedance Fault (b) *CumSum* of fault current of segment 1 (c) Wavelet of fault current of segment 1 (d) Segment 1 current before and after fault clearing.

In this case, the time taken to detect the fault is independent of the fault impedance. The fault is cleared by power switches within 150μs.

7.3 FAULT LOCALISATION IN THE SEGMENT

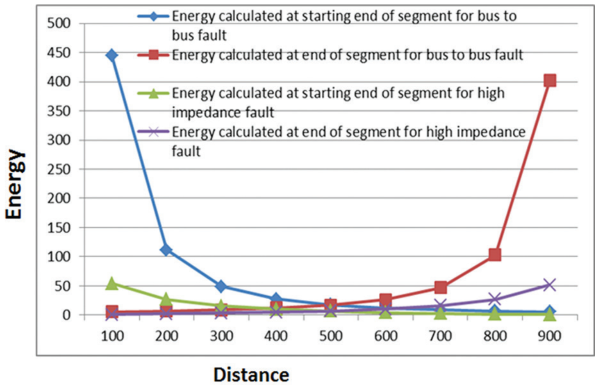


Fig. 19. Energy concentration vs distance in faulty segment due to bus to bus fault and high impedance fault

To check the effectiveness of algorithm for fault location identification faults are initialized after every 100m in each segment from the starting point to the end of the segment. Current is being measured at both the end of the segment and the change in energy is calculated from the wavelet high pass coefficient. The calculated energy from the starting of segment 2 to the endpoint of segment 2 for the bus to bus fault and high impedance fault is shown in figure 19, after initializing the fault in segment 2. If a bus to bus fault occurred at a point 100m from the starting point of segment 2, energy value is 450 J at starting point and 6 J at the end of segment 2. For the bus to bus fault at a point 900 m from the endpoint of segment 2 has energy value 400 J at the endpoint and 5 J at starting point of segment 2. Similarly, a change in energy values can be observed for high impedance faults. The graph is plotted for the wavelet energy changes concerning the distance in meters for segment 2. This change of energy is utilized to distinguish the location of the fault.

The developed simulation model is tested with other fault detection methods like differential fault current, over current, di/dt, and wavelet transform. Figure 20 shows the comparison of the proposed method with other fault detection methods in terms of selectivity, reliability, and accuracy. The accuracy is the ratio of number trail in which fault detect correctly to total number of trails. Selectivity is the property of protection scheme to respond to faulty zone and not others. Reliability of protection scheme is to performing its function adequately for the intended period of time under specified operating conditions. The comparison shows that the proposed method gives an accuracy of 98.72%, whereas the accuracy achieved by other methods are Wavelet transform of 92.43%, di/dt of 90.12%, overcurrent protection of 80.44%, and differential protection, 88.25%. Figure 20 also shows proposed method comparatively has the highest reliability of 99.01% and selectivity of 96.08%.

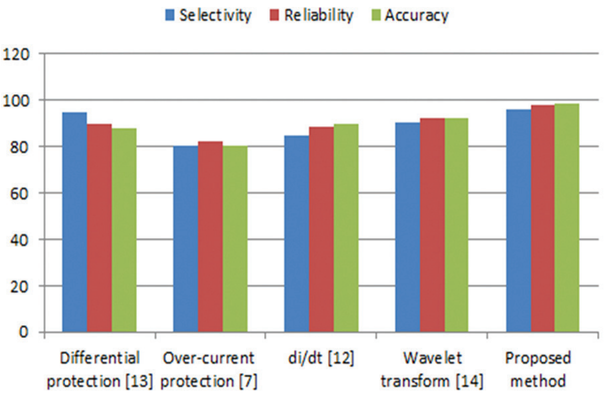


Fig. 20. Selectivity, reliability and accuracy comparison of the proposed method with other methods

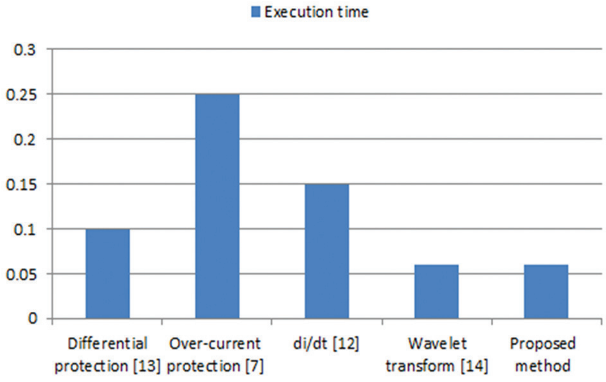


Fig. 21. Comparison of Execution time of the proposed method with other methods

Figure 21 shows the comparison of the proposed method with the other methods in terms of execution time. It shows that the proposed method takes lesser time to execute than wavelet transform, di/dt, overcurrent protection, and differential protection. Furthermore, the proposed method has a high sensitivity to high impedance fault comparatively overcurrent and non-unit protection method. The proposed technique does not have to rely on communication between protective relays on both sides of the protected zone for fault detection as differential protection. The artificial neural network-based method uses two networks for fault detection and fault localization and has a complex classifier structure which leads to long training and detection time. Comparatively, the proposed method has high intelligent fault detection capability, robust and very fast selective fault isolation capability.

8. Conclusion

An efficient protection scheme using *CumSum* and Wavelet transform is proposed and implemented in a Ring-type LVDC microgrid system. The fault is being detected from the fault current average using *CumSum* and fault features extracted using Wavelet transform. Depending on the severity of the fault, the *CumSum* of even a healthy section crosses the set threshold value which may lead to the maloperation of the healthy system. However, the Wavelet transform can identify the faulty segment irrespective of the severity and location of the fault. During sudden changes in load, *CumSum* value is not changing much whereas wavelet coefficient goes above the set threshold value, indicating even load change as a fault. The hybrid model works well under fault conditions and dynamic conditions and can distinguish between dynamic operation and a fault in the DC microgrid. The threshold values for *CumSum* and Wavelet transform are calculated analytically and the same is used in the algorithm. Energy per cycle is determined from the wavelet coefficient and is used to find the fault location, accordingly taking out the requirement for additional hardware to distinguish the fault area. From the comparative study, it is shown that the proposed method has higher accuracy 98.72%, higher selectivity 96.08%, higher reliability 99.01%, and lower execution rate in comparison with the existing fault detection methods in the discrimination of fault and system dynamic situations. The proposed methodology is robust, highly sensitive to high impedance fault, does not require communication between protection devices, and has fast selective fault isolation capability. The research can be additionally extended to optimum load sharing and stability study. The methodology will be more advantageous if it can handle system dynamics to maintain the stability of the system from the span of occurrence of fault to its isolation. A more detailed control approach can be evolved to make the system stable and reliable.

9. REFERENCES

[1] Y. Ito, Z. Yang, A. Hirofumi, "DC microgrid based distribution power generation system", Proceedings of the 4th International Power Electronics and Motion Control Conference, 2004.

[2] S. Augustine et al. ″DC Microgrid Protection: Review and Challenges", Technical report, Sandia National Laboratory, Albuquerque, NM, USA, 2018.

[3] A. Chandra, G. K. Singh, V. Pant, "Protection techniques for DC microgrid-A review", Electric Power Systems Research, Vol. 187, 2020, p. 106439.

[4] S. Som, S. R. Samantaray, "Efficient protection scheme for low-voltage DC micro-grid", IET Generation, Transmission & Distribution, Vol.12, No. 13, 2018, pp. 3322-3329.

[5] D. Kumar, F. Zare, A. Ghosh, "DC microgrid technology: system architectures, AC grid interfaces, grounding schemes, power quality, communication networks, applications, and standardizations aspects", IEEE Access, Vol. 5, 2017, pp. 12230-12256.

[6] R. Kamel, K. Nagasaka, "Effect of load type on standalone micro grid fault performance", Applied energy, Vol. 160, 2015, pp. 532-540.

[7] P. Mahat et al. "A simple adaptive overcurrent protection of distribution systems with distributed generation", IEEE Transactions on Smart Grid, Vol. 2, No. 3, 2011, pp. 428-437.

[8] R. Mohanty, S. M. Balaji, A. K. Pradhan, "An accurate noniterative fault-location technique for low-voltage DC microgrid", IEEE Transactions on Power Delivery, Vol. 31, No. 2, 2015, pp. 475-481.

[9] J. Wang et al. "Fast fault selection and location for a marine power system using system power converters, and active impedance estimation", Proceedings of the 4th IET Conference on Power Electronics, Machines and Drives, 2008.

[10] F. Wilches-Bernal et al. "A Survey of Traveling Wave Protection Schemes in Electric Power Systems", IEEE Access, Vol. 9, 2021, pp. 72949-72969.

[11] S. S. Balasreedharan, S. Thangavel, "An adaptive fault identification scheme for DC microgrid using event based classification", Proceedings of the 3rd International Conference on Advanced Computing and Communication Systems, 2016.

[12] A. Meghwani, S. Srivastava, S. Chakrabarti, "A new protection scheme for DC microgrid using line current derivative", Proceedings of the IEEE/PES General Meeting, 2015, pp.1-5.

[13] D. Salomonsson, L. Soder, A. Sannino, "Protection of low-voltage DC microgrids", IEEE Transactions on Power Delivery, Vol. 24, No. 3, 2009, p. 1045.

[14] D. K. J. S. Jayamaha, N. W. A. Lidula, A. D. Rajapakse, "Wavelet-multi resolution analysis based ANN architecture for fault detection and localization in DC microgrids", IEEE Access, Vol. 7, 2019, pp. 145371-145384.

[15] D. Patil, S. Bindu, S. Thale, "Deep neural netwok enable fault deecting in LVDC microgrid using empirical mode decompositin", International Journal of Advance Technology and Engineering Eploration, Vol. 9, No. 87, 2022; pp. 200-215.

[16] J.-D. Park, J. Candelaria, "Fault detection and isolation in low-voltage DC-bus microgrid system", IEEE transactions on power delivery, Vol. 28, No. 2, 2013, pp. 779-787.

[17] Y. Wang, Z. Zhang, Y. Fu, Y. Hei, X. Zhang, "Pole-to-ground fault analysis in transmission line of DC grids based on VSC", Proceedings of the IEEE 8th International Power Electronics and Motion Control Conference, 2016.

[18] A. Meghwani, S. C. Srivastava, S. Chakrabarti, "A non-unit protection scheme for DC microgrid based on local measurements", IEEE Transactions on Power Delivery, Vol. 32, No. 1, 2016, pp. 172-181.

[19] S. Fletcher et al. "High-speed differential protection for smart DC distribution systems", IEEE Transactions on Smart Grid, Vol. 5, No. 5, 2014, pp. 2610-2617.

[20] D. Patil, S. Bindu, "Real time protection technique for DC microgrid using local measurements", Proceedings of the Technologies for Smart-City Energy Security and Power Conference, 2018.

[21] S. Beheshtaein et al. "DC microgrid protection: A comprehensive review", IEEE Journal of Emerging and Selected Topics in Power Electronics, 2019. (in Press)

[22] R. Polikar, "The Wavelet Tutorial Second Edition Part I", https://ccrma.stanford.edu/~unjung/mylec/WTpart1.html (accessed: 2021)

[23] D. Obertson, O. Camps, J. Mayer, W. Gish, 'Wavelets and electromagnetic power system transients', IEEE Transactions on Power Delivery, Vol. 11, No. 2, 1996, pp. 1050-1058

[24] D. Patil, G. Pushkar, V. Patwardhan, M. Gadre, "On the design of FIR wavelet filter banks using factorization of a halfband polynomial", IEEE Signal Processing Letters, Vol. 15, 2008, pp. 485-488.

[25] K. De Kerf et al. "Wavelet-based protection strategy for DC faults in multi-terminal VSC HVDC systems", IET Generation, Transmission & Distribution, Vol. 5, No. 4, 2011, pp. 496-503.

[26] U. Maqbool, U. A. Khan, "Wavelet Based Feature Analysis of Fault Signals in a Microgrid", Proceedings of the International Conference on Power Generation Systems and Renewable Energy Technologies, 2018.

[27] J.-D. Park, J. Candelaria, L. Ma, K. Dunn, "DC ring-bus microgrid fault protection and identification of fault location", IEEE transactions on Power delivery, Vol. 28, No. 4, 2013, pp. 2574-2584.

[28] C. Patil et al. "A novel protection scheme for DC microgrid with hierarchical control", Proceedings of the IEEE International Conference on Smart Energy Grid Engineering, 2017.

Classification of Healthcare Service Reviews with Sentiment Analysis to Refine User Satisfaction

Review Paper

Khai Herng Leong
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
khaiherngleong98@gmail.com

Dahlila Putri Dahnil
Centre for Software Technology and Management (SOFTAM),
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
dahlilaputri@ukm.edu.my

Abstract *– In natural language processing, sentiment analysis determines the polarity of a message based on lexical emotion. This technique is utilized intensively in service sectors to study the level of consumer satisfaction. However, the healthcare service field lacks such practice to detail responses in existing feedback systems. A proposed application which implements sentiment analysis is developed for improvement. User reviews are classified according to their word influences, namely positive, negative and neutral states. In addition, topic modelling is included to organize them in several service themes. A graphical user interface, GUI which records the analytical results is presented to users for interaction. This approach does not only benefit patients to choose their desired medical centres, but also healthcare management who wish to enhance their service quality.*

Keywords*: Healthcare Service Review System, Natural Language Processing, Sentiment Analysis, Topic Modelling, Web Scraping*

1. INTRODUCTION

Healthcare preserves human health through prevention, detection, as well as treatment of disabilities, illness, injuries and mental. Medical tourism to Malaysia is becoming popular because of high quality healthcare service with low cost compared to other countries in Asia region [1].

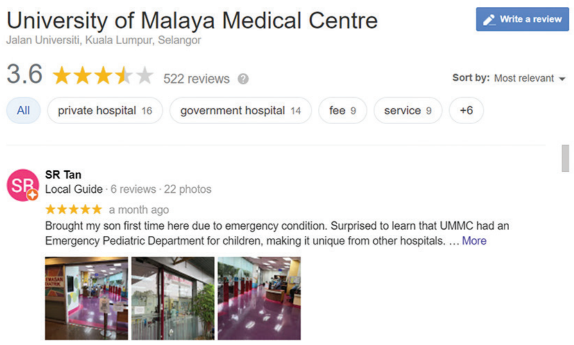
We often hear about healthcare service issues of certain medical centres in our community. Most of the customers express their opinion about the medical services they received to closest people, while some may give feedback via online to share their experiences. These responses are very useful for patients to choose a medical centre based on the testimonies. Customers experiences and testimonies are able to help other patients in the selection of medical centres. Patients can compare the satisfaction details in choosing which medical centre they would like to seek treatment from. Furthermore, medical centre management can refer to this application for service improvement. Thus, a comprehensive healthcare service feedback system is required to classify the reviews.

A sentiment analysis algorithm is developed to provide classification to the user reviews. This technique emphasizes word emotion to determine the polarity of a sentence. Scattered responses can be arranged neatly based on evaluation parameters. All functionalities are implemented to facilitate the process of referral, comparison and selection of medical centres.

2. LITERATURE REVIEW

Internet feedback system collects ratings and reviews from users as references for the community. The former acts as a Likert scale which comprises five or ten points. The higher the rating, the better the reputation. Next, the latter states the aspects of service quality, such that we can comprehend its general description. The existing systems which support healthcare service are Google review, Lyfboat and Yelp.

Google review is a popular feedback system based on world location, where users can rate and leave responses for any services. The application extracts several keywords according to mentioned frequency to present information more effectively. All reviews can be arranged in four order types, namely most relevant, newest, highest and lowest rating. A language translation tool, Google Translate is utilized to interpret foreign comments. Fig. 1 shows the interface of the Google review.

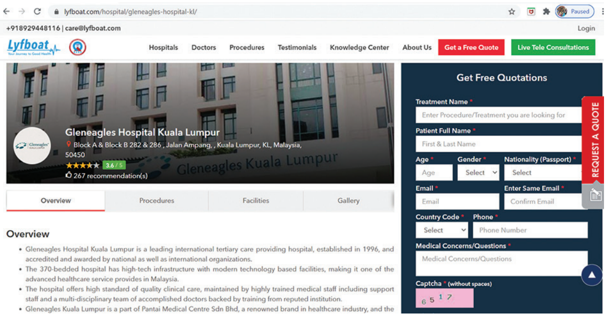


University of Malaya Medical Centre
Jalan Universiti, Kuala Lumpur, Selangor
3.6 ★★★★☆ 522 reviews

SR Tan
Local Guide · 6 reviews · 22 photos
★★★★★ a month ago
Brought my son first time here due to emergency condition. Surprised to learn that UMMC had an Emergency Pediatric Department for children, making it unique from other hospitals. …

Fig. 1. The Interface of Google Review

Lyfboat is an international healthcare website which provides search and query functionalities for treatment procedures, doctor and hospital information. The latter has neat details including centre excellence, medical infrastructure and transportation facility. Nonetheless, at most 12 well known Malaysian hospitals are recorded in the database. Besides, star rating is the only feedback channel. The interface is shown in Fig. 2.



lyfboat.com/hospital/gleneagles-hospital-kl/
+918929448116 | care@lyfboat.com

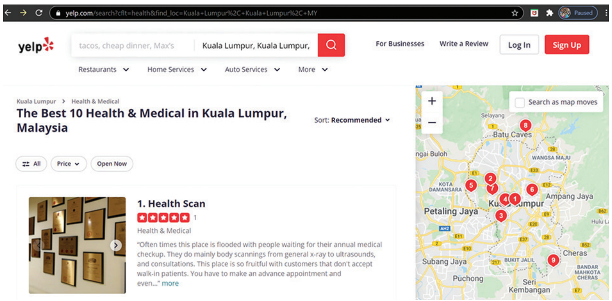
Gleneagles Hospital Kuala Lumpur
Block A & Block B 282 & 286 , Jalan Ampang, , Kuala Lumpur, KL, Malaysia, 50450
★★★★★ 3.6 / 5
267 recommendation(s)

Overview
- Gleneagles Hospital Kuala Lumpur is a leading international tertiary care providing hospital, established in 1996, and accredited and awarded by national as well as international organizations.
- The 370-bedded hospital has high-tech infrastructure with modern technology based facilities, making it one of the advanced healthcare service provides in Malaysia.
- The hospital offers high standard of quality clinical care, maintained by highly trained medical staff including support staff and a multi-disciplinary team of accomplished doctors backed by training from reputed institution.
- Gleneagles Kuala Lumpur is a part of Pantai Medical Centre Sdn Bhd, a renowned brand in healthcare industry, and the

Get Free Quotations
Treatment Name * Enter Procedure/Treatment you are looking for
Patient Full Name * First & Last Name
Age * Age
Gender *
Nationality (Passport) *
Email * Email
Enter Same Email * Confirm Email
Country Code *
Phone * Phone Number
Medical Concerns/Questions * Medical Concerns/Questions
Captcha * (without spaces) 6517

Fig. 2. Lyfboat interface

Yelp is a mobile application and business service website which covers Malaysia and 31 other countries. It emphasizes comprehensive search and comparison between premises in terms of distance, evaluation, price and others. Medical centres are included to provide healthcare service references. However, there are less reviews available because it is not a common practice among Malaysians. The interface of the Yelp website is shown in Fig. 3.



yelp.com/search?cflt=health&find_loc=Kuala+Lumpur%2C+Kuala+Lumpur%2C+MY

Kuala Lumpur > Health & Medical
The Best 10 Health & Medical in Kuala Lumpur, Malaysia
Sort: Recommended

1. Health Scan
★★★★★ 1
Health & Medical
"Often times this place is flooded with people waiting for their annual medical checkup. They do mainly body scannings from general x-ray to ultrasounds, and consultations. This place is so fruitful with customers that don't accept walk-in patients. You have to make an advance appointment and even..." more

Fig. 3. Yelp interface

Sentiment analysis is used for biometrics, computational linguistics, natural language processing, NLP, as well as text analysis in distinguishing, extracting, quantifying and studying subjective information. It is divided to two approaches, namely lexicon and machine learning. The latter consists of classification algorithm which trains dataset samples to predict the polarity of other documents [2]. Linear regression, Naive Bayes and support vector machine are some of the models which are often applied in this field. On the other hand, lexicon method estimates polarity scores based on word matches and relevant sentimental lexicons [3]. The score ranges from -1 to 1, where below shows how to label it:

- Equal or more than 0.05 is positive state.
- Equal or less than -0.05 is negative state.
- Between -0.05 and 0.05 is neutral state.

Hence, sentiment analysis identifies whether a sentence is positive, negative or neutral state. The accuracy of this system depends on the efficiency of human judgement. An algorithm which achieves a 70% validity level is considered to have good machine intelligence as we agree about any events in an average of 80% only.

In terms of these approaches, lexicon method is more superior than statistically trained classifiers [8]. Lexicon extension with linguistic information improves system durability. The correlation between frequency of keyword and overall rating of text is stated clearly, which guarantees the quality in generating lexicons.

3. THE PROPOSED SYSTEM

Five hospitals and clinics selected in the Google review feedback obtained from the website for each hospital [9-13] The processing from the feedback to the GUI is presented in Fig. 4. Initially, the data is scraped and stored in an integrated development environment, IDE. Topic modelling was conducted in each response to identify service areas in the information classification. Next, sentiment analysis was implemented to obtain polarity scores based on the probability of the appearance of character emoticons [4]. This index is capable of calculating the cumulative average of the ratings and determining the nature of the polarity to classify the responses.

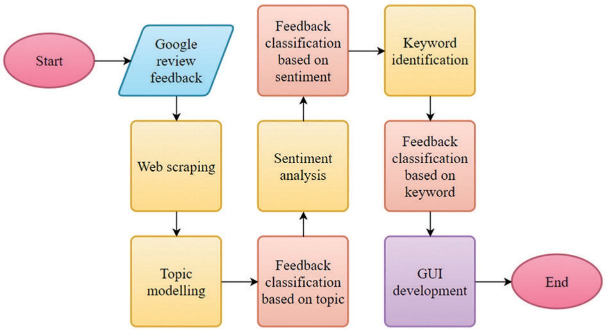
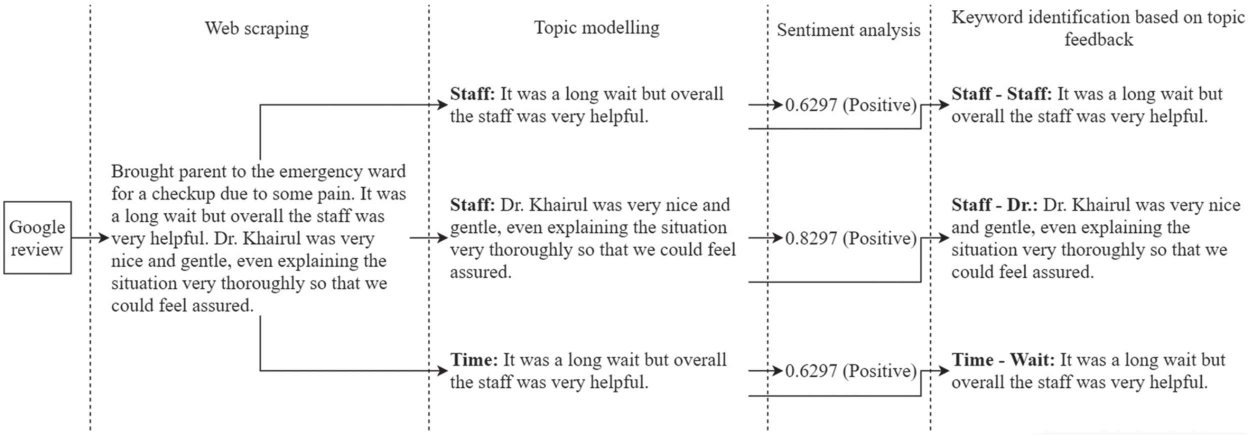

Start
Google review feedback
Web scraping
Topic modelling
Feedback classification based on topic
Sentiment analysis
Feedback classification based on sentiment
Keyword identification
Feedback classification based on keyword
GUI development
End

Fig. 4. Architectural design of algorithm

Frequently used words are also identified to indicate reviews that have these words. Finally, a GUI is built to deliver all processed health service feedback information to users. The approach and text examples are shown in Fig. 5.



Brought parent to the emergency ward for a checkup due to some pain. It was a long wait but overall the staff was very helpful. Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.

Staff: It was a long wait but overall the staff was very helpful.

Staff: Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.

Time: It was a long wait but overall the staff was very helpful.

0.6297 (Positive)

0.8297 (Positive)

0.6297 (Positive)

Staff - Staff: It was a long wait but overall the staff was very helpful.

Staff - Dr.: Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.

Time - Wait: It was a long wait but overall the staff was very helpful.

Fig. 5. Algorithm and text processing steps

3.1. WEB SCRAPING

A list of feedback, star ratings and premise information in a Google review is required to run this project. In addition to copy and paste, web scraping is an alternative and efficient method of collecting duplicate data. The mechanism of the technique is to crawl hypertext markup language, HTML and extract the desired information through programming. The HTML parser activates the application programming interface, API to access website content [6]. The Beautiful Soup and Selenium WebDriver modules were used throughout this procedure. The flow of data from a Google review to be recorded in a spreadsheet and text file is presented in Fig. 6.

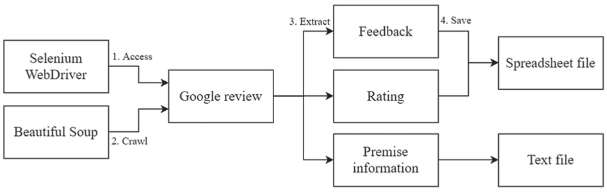


Fig. 6. Process of web scraping

Web scraping involves four steps, namely access, crawl, extract and save, with the following essence:

1. **Access**: WebDriver is programmed to emulate users browsing Google reviews in reading medical centre information.
2. **Crawl**: Beautiful Soup crawls Google review HTML to find the data it needs.
3. **Extract**: Feedback lists, star ratings and premise information were extracted.
4. **Save**: This information is stored in a spreadsheet and text file for use in the next stage.

The five hospitals and clinics surveyed are as follows:

- Subang Jaya Medical Centre, SJMC.
- Ara Damansara Medical Centre, ADMC.
- Tung Shin Hospital.
- Universiti Kebangsaan Malaysia Medical Centre, UKMMC.
- The KL Sky Clinic.

3.2. TOPIC MODELLING

Topic modelling identifies latent themes that potentially describe a piece of text [7]. Latent Dirichlet allocation, LDA are among the popular unsupervised machine learning algorithms in this approach. The technique detracts from previous Dirichlets for distributing topics and words, as well as avoiding overfitting effects [5]. This model assumes all documents are generated through a statistical generation process, meaning they contain a number of speculative titles from a list of related keywords. The flow of topic modeling in formulating the feedback theme, in which Subang Jaya Medical Centre (SJMC) serves as a modeling example is presented in Fig. 7.

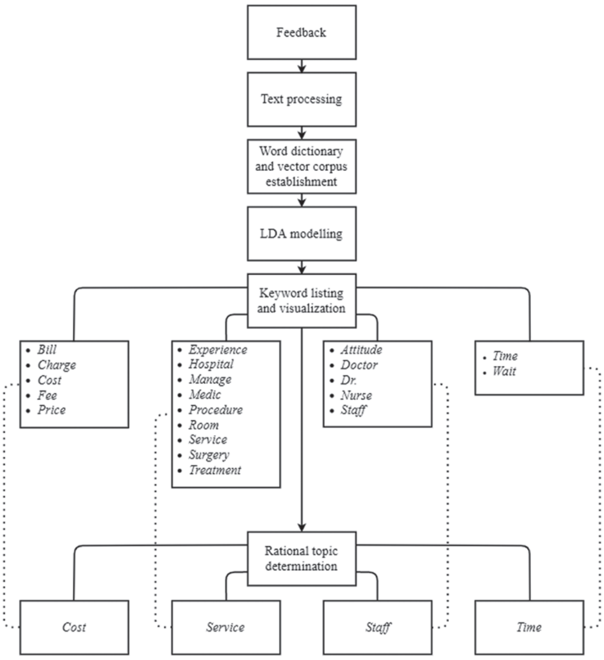


Fig. 7. Process of topic modelling with SJMC as example

This procedure was also carried out in response to ADMC, Tung Shin Hospital, Universiti Kebangsaan Malaysia Medical Centre (UKMMC) and The KL Sky Clinic.

First, text processing is implemented to restore word structure in constructing vector dictionaries and corpora. Both of these matrices were modelled with the LDA algorithm to generate interactive lists and graphs of keywords representing various types of themes.

The role of the developer is to identify as many as four service topics from a particular word that have a semantic relationship based on the above framework. For example, 'bill', 'charge', 'cost', 'fee' and 'price' can interpret 'Cost'. The list of titles in each medical centre is as follows:

- **SJMC**: 'Cost', 'Service', 'Staff' and 'Time'.
- **ADMC**: 'Park', 'Service', 'Staff' and 'Time'.
- **Tung Shin Hospital**: 'Cost', 'Service', 'Staff' and 'Time'.
- **UKMMC**: 'Park', 'Service', 'Staff' and 'Time'.
- **The KL Sky Clinic**: 'Service' and 'Staff'.

Each feedback may contain at least one theme that can be classified. The method of classification according to keywords and topics is presented in Fig. 8.

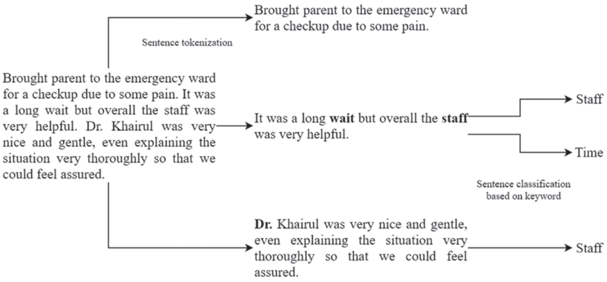

Brought parent to the emergency ward for a checkup due to some pain.
Sentence tokenization
Brought parent to the emergency ward for a checkup due to some pain. It was a long wait but overall the staff was very helpful. Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.
It was a long wait but overall the staff was very helpful.
Staff
Time
Sentence classification based on keyword
Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.
Staff

Fig. 8. Feedback classification based on keyword and topic

Response tokenization is performed to isolate verses. If it has a themed word, the sentence is categorized in a related topic, as in the illustration above. 'Dr.' and 'staff' represent 'Staff', while 'wait' is 'Time'. The process was also conducted on other reviews in five medical centres.

3.3. SENTIMENT ANALYSIS

Sentiment analysis is the emotional research of texts in positive, negative and neutral, where this status is known as the nature of polarity. Essentially, the three traits signify cheerful, hateful and moderate feelings respectively.

In this project, the Valence Aware Dictionary for Sentiment Reasoning model, VADER is implemented to determine sentence sentiment and calculate the average polarity of service topics through scoring. The score range is from -1 to 1, where equal to or greater than 0.05 signifies positive, equal to or minus -0.05 signifies negative, and between -0.05 and 0.05 signifies neutral. Table 1 shows the nature of topic polarity in each medical centre.

According to the table, the reputations of ADMC and The KL Sky Clinic are satisfactory, while SJMC, Tung Shin Hospital and UKMMC are modest.

Table 1. Polarity of service topics

Medical centre	Service topic	Average mark	Polarity	Rating
SJMC	Cost	-0.03	Neutral	3
	Service	0.03	Neutral	3
	Staff	0.11	Positive	3
	Time	-0.03	Neutral	3
ADMC	Park	0.39	Positive	4
	Service	0.15	Positive	3
	Staff	0.21	Positive	4
	Time	0.03	Neutral	3
Tung Shin Hospital	Cost	0	Neutral	3
	Service	0	Neutral	3
	Staff	-0.04	Neutral	3
	Time	-0.11	Negative	3
UKMMC	Park	-0.04	Neutral	3
	Service	-0.03	Neutral	3
	Staff	0.01	Neutral	3
	Time	-0.05	Negative	3
The KL Sky Clinic	Service	0.44	Positive	4
	Staff	0.46	Positive	4

3.4. KEYWORD IDENTIFICATION

Before calculating word frequency, text processing should be performed to restore formatting and get rid of less meaningful keywords. Examples of sentences used in the demonstration are as follows:

The 3 nurses are good,
can better!

A description of this procedure is attached with the text output in Table 2.

Table 2. Text processing

Step	Process	Output
1	Replace line breaks with spaces.	The 3 nurses are good, can better!
2	Remove tabs.	The 3 nurses are good, can better!
3	Replace '&amp' with '&'.	The 3 nurses are good, can better!
4	Remove '(Translated by Google)', if any.	The 3 nurses are good, can better!
5	Remove the original review not in English, if any.	The 3 nurses are good, can better!
6	Remove accented letters.	The 3 nurses are good, can better!
7	Remove digits.	The nurses are good, can better!
8	Remove punctuation.	The nurses are good can better
9	Convert uppercase to lowercase.	the nurses are good can better
10	Remove stopwords.	nurses good better
11	Perform word lemmatization.	nurse good good

12	Perform word stemming.	nurs good good
13	Perform word tokenization.	nurs, good, good
14	Calculate the cumulative frequency of words.	nurs (1), good (2)

A word cloud consist of a collection of keywords and the size of each keyword determines their dominance in a content. Fig. 9 shows the collection of keywords and the significant keyword is determined based on its size.

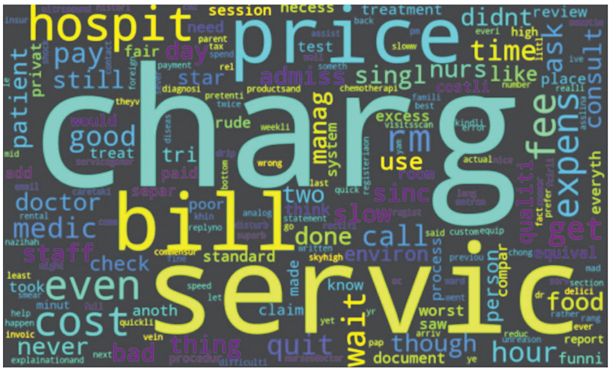


Fig. 9. Word cloud

The cumulative number of words was calculated to determine the five key keywords that were frequently specified in the response to each service topic. All reviews that have the word are organized in their respective groups. Table 3 presents this list of contents.

Table 3. Most mentioned keywords of service topics

Medical centre	Service topic	Keyword
SJMC	Cost	Charge
		Price
		Service
		Bill
		Hospital
	Service	Medic
		Service
		Hospital
		Doctor
		Experience
	Staff	Staff
		Dr.
		Nurse
		Doctor
		Hospital
	Time	Wait
		Time
		Hospital
		Doctor
		Hour
ADMC	Park	Park
		Easy
		Spacious
		Clean
		Hospital
	Service	Service
		Medic
		Hospital
		Doctor
		Room
	Staff	Nurse
		Friendly
		Doctor
		Dr.
		Staff
	Time	Time
		Wait
		Hour
		Doctor
		Appoint
Tung Shin Hospital	Cost	Service
		Price
		Hospital
		Doctor
		Ask
	Service	Good
		Service
		Hospital
		Medic
		Doctor
	Staff	Doctor
		Nurse
		Dr.
		Staff
		Hospital
	Time	Ask
		Wait
		Time
		Doctor
		Hour
UKMMC	Park	Park
		Hospital
		Fee
		Rate
		Expensive
	Service	Hospital
		Service
		Patient
		Doctor
		Good

UKMMC	Staff	Nurse
		Staff
		Patient
		Doctor
		Time
	Time	Time
		Wait
		Doctor
		Patient
		Hospital
The KL Sky Clinic	Service	Treatment
		Clinic
		Service
		Medic
		Dr.
	Staff	Dr.
		Doctor
		Roland
		Treatment
		Friendly

4. MODEL VALIDATION

In terms of sentiment analysis, the motive of this section is to investigate how effective the VADER model is in classifying the nature of polarity in line with human feelings. The sentiment evaluation process of the technique as well as Google review users is presented in Fig. 10 and Fig. 11.

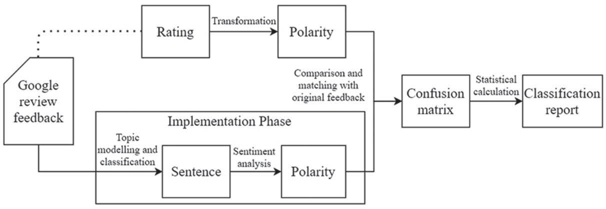


Fig. 10. Testing process of sentiment analysis

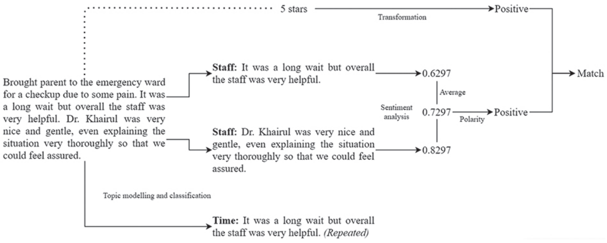

5 stars
Transformation
Positive
Brought parent to the emergency ward for a checkup due to some pain. It was a long wait but overall the staff was very helpful. Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.
Staff: It was a long wait but overall the staff was very helpful.
0.6297
Average
Sentiment analysis
0.7297
Polarity
Positive
Match
Staff: Dr. Khairul was very nice and gentle, even explaining the situation very thoroughly so that we could feel assured.
0.8297
Topic modelling and classification
Time: It was a long wait but overall the staff was very helpful. (Repeated)

Fig. 11. Comparison and matching process of polarity

Google review star feedback and ratings are provided by users with a good visit experience. Star points range from 1 to 5, where data transformations are also carried out to categorize them in positive, negative and neutral properties as comparative benchmarks. The details of the classification are as follows:

- 1 and 2 stars are made as negative state.
- 3 stars is made as neutral state.
- 4 and 5 stars are made as positive state.

Each response may contain themed sentences of several topics that can be identified through topic analysis. In the implementation phase, they were extracted and classified in related topics, then conducted sentiment analysis to obtain polarity scores. Further, the mean score was calculated to determine the nature of the overall polarity which also consisted of positive, negative and neutral. It was observed that sentences repeated in other topics were ignored to avoid biased decisions.

In the example above, the Google review feedback is rated 5 stars, which is a positive attribute. After conducting topic analysis and classification, the sentences were divided into the themes which are 'Staff' and 'Time' respectively. The result of the process is presented in a standalone system with a GUI to enable users to view the feedback shown in Figure 12. Sentiment analysis was performed to obtain their polarity scores, namely 0.6297 and 0.8297. Average scores were also calculated, where repetitive sentences were excluded. The result is 0.7297 which indicates a positive status, which is similar to the original sentiment.

The purpose of this process is to compare machine sentiment with that of humans to investigate the level of accuracy. Matches between these two data were recorded in a confusion matrix, as in Table 4.

684 feedback fractures were compared. Of the 227 negative traits, 116 neutrals and 341 positive of origin, VADER attempted to determine 204, 17 and 249 respectively. In addition, of the 338 negative traits, 51 were neutral and 295 positive predictions, 204, 17 and 249 were corresponding to the original sentiments, respectively. Indeed, a classification report can be tabulated with this data.

The report discusses four criteria that describe the classification statistics, namely accuracy, retrieval, F-score and accuracy. Table 5 presents the calculation results according to the confusion matrix.

Table 5. Classification report

	Precision	Recall	F-score
Negative	0.9	0.6	0.72
Neutral	0.15	0.33	0.2
Positive	0.73	0.84	0.78
Accuracy		0.69	

Precision checks the efficiency of the classifier in predicting the original valuation. Recall emphasizes the accuracy of the classifier in predicting the original valuation. Next, F-score tests the accuracy and recall performance. Finally, accuracy determines the overall percentage of sentiment that is correctly predicted by the comparison volume.

Table 4. Confusion matrix of polarity

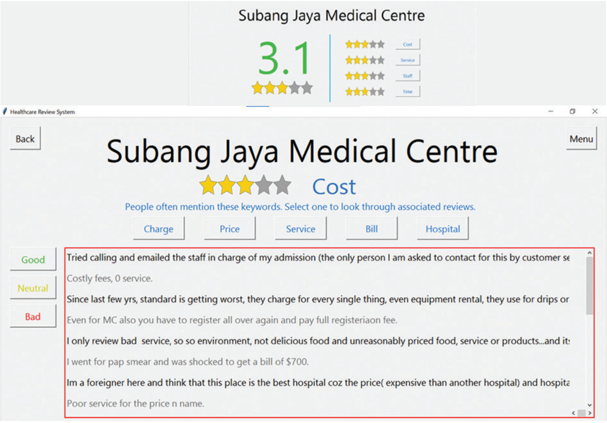
		Original			Total predicted sentiment
		Negative	Neutral	Positive	
Prediction	Negative	204	63	71	338
	Neutral	13	17	21	51
	Positive	10	36	249	295
Total original sentiment		227	116	341	684

A level of accuracy that reaches 69% means that the VADER model is good at identifying user feedback feelings. In terms of the F-score, the neutral status underperformed because the response may contain a mixture of positive and negative sentences with extreme polarity. In contrast, the positive and negative traits showed very good results throughout the classification procedure. Negative sentiments with the highest percentage of precision indicated that the model was able to label nine out of ten reviews as bad, while positive sentiments with the highest recall values referred to the algorithm performing in determining good responses with minimum error rates.

5. CONCLUSION

Health services are a primary need because of their role in healing and saving human beings. Therefore, the quality of services must be taken care of to guarantee the universal interest. A comprehensive feedback system has been built to try to sustain this mission.

Compared to existing systems, this new application recognizes sentiment analysis techniques that classify feedback into three polarities, namely positive, negative and neutral. In addition, topic analysis and key word determination were also implemented to further detail user feelings. Such advantages make it easy for the patient to peruse the details of the review and select the desired medical centre. Healthcare management can also refer to this system to improve the quality of their services.


Subang Jaya Medical Centre
3.1
Cost
Service
Staff
Time
Healthcare Review System
Back
Menu
Subang Jaya Medical Centre
Cost
People often mention these keywords. Select one to look through associated reviews.
Charge
Price
Service
Bill
Hospital
Good
Neutral
Bad

Tried calling and emailed the staff in charge of my admission (the only person I am asked to contact for this by customer se

Costly fees, 0 service.

Since last few yrs, standard is getting worst, they charge for every single thing, even equipment rental, they use for drips or

Even for MC also you have to register all over again and pay full registeriaon fee.

I only review bad service, so so environment, not delicious food and unreasonably priced food, service or products...and it:

I went for pap smear and was shocked to get a bill of $700.

Im a foreigner here and think that this place is the best hospital coz the price(expensive than another hospital) and hospita

Poor service for the price n name.

Fig. 12. The GUI to show the result after topic analysis and classification.

6. ACKNOWLEDGMENT

The research is funded by Universiti Kebangsaan Malaysia (UKM) research project DCP-2018-001/2 under Research Centre for Software Technology and Management (SOFTAM), www.ftsm.ukm.my/softam, Faculty of Information Science and Technology.

7. REFERENCE

[1] International Living, Healthcare in Malaysia. https://internationalliving.com/countries/malaysia/healthcare-in-malaysia (accessed: 2021)

[2] M. Birjali, A. Beni-Hssane, M. Erritali, “Machine Learning and Semantic Sentiment Analysis Based Algorithms for Suicide Sentiment Prediction in Social Networks”, Procedia Computer Science, Vol. 113, 2017, pp. 65-72.

[3] O. Araque, G. Zhu, C. A. Iglesias, “A Semantic Similarity-based Perspective of Affect Lexicons for Sentiment Analysis”, Knowledge-Based Systems, Vol. 165, 2019, pp. 346-359.

[4] K. Utsu, J. Saito, O. Uchida, “Sentiment Polarity Estimation of Emoticons by Polarity Scoring of Character Components”, Proceedings of the IEEE Region Ten Symposium, Sydney, NSW, Australia, 4-6 July 2018.

[5] R. Annisa, I. Surjandari, Zulkarnain, “Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation”, Proceedia Computer Science, Vol. 161, 2019, pp. 739-746.

[6] V. Singrodia, A. Mitra, S. Paul, “A Review on Web Scrapping and Its Applications”, 2019 Proceedings of the International Conference on Computer Communication and Informatics, Coimbatore, India, 23-25 January 2019.

[7] S. Kim, H. Park, J. Lee, "Word2vec-based Latent Semantic Analysis (W2V-LSA) for Topic Modeling: A Study on Blockchain Technology Trend Analysis", Expert Systems With Applications, Vol. 152, 2020, p. 113401.

[8] D. Grabner, M. Zanker, G. Fliedl, M. Fuchs, "Classification of Customer Reviews Based on Sentiment Analysis", Proceedings of the 19th Conference on Information and Communication Technologies in Tourism, Helsingborg, Sweden, 25-27 January 2012, pp. 460-470.

[9] Pusat Perubatan Subang Jaya, Subang Jaya Medical Centre, https://www.google.com/maps/place/Subang+Jaya+Medical+Centre+(SJMC)/@3.0765703,101.5909797,15.92z/data=!4m7!3m6!1s0x31cc4c60badceb4f:0xa2a80452021765a6!8m2!3d3.079771!4d101.5938418!9m1!1b1

[10] Pusat Perubatan Ara Damansara, Ara Damansara Medical Centre, https://www.google.com/maps/place/Ara+Damansara+Medical+Centre/@3.1151723,101.5627106,17z/data=!4m7!3m6!1s0x31cc4e7f4b90f90d:0x34c9559e5d246762!8m2!3d3.1151669!4d101.5648993!9m1!1b1 (accessed: 2021)

[11] Hospital Tung Shin, https://www.google.com/maps/place/Tung+Shin+Hospital/@3.1459617,101.7016377,17z/data=!4m11!1m2!2m1!1stung+shin!3m7!1s0x31cc49d42ad29b1d:0xcbbb8b9f09c732e6!8m2!3d3.1464757!4d101.7040118!9m1!1b1!15sCgl0dW5nIHNoaW5aCyIJdHVuZyBzaGlukgEIaG9zcGl0YWywAQA (accessed: 2021)

[12] Pusat Perubatan Universiti Kebangsaan, Malaysia, https://www.google.com/maps/place/Pusat+Perubatan+Universiti+Kebangsaan+Malaysia/@3.0992724,101.7232131,17z/data=!4m7!3m6!1s0x31cc35e90db3a4ad:0xaf6721771e83594a!8m2!3d3.099267!4d101.7254018!9m1!1b1 (accessed: 2021)

[13] The KL Sky Clinic, https://www.google.com/maps/place/The+KL+Sky+Clinic/@3.1536606,101.7079711,17z/data=!4m7!3m6!1s0x31cc37d4514100eb:0xb390c52a4cc315e3!8m2!3d3.1536552!4d101.7101598!9m1!1b1 (accessed: 2021)

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia.

About this Journal

The International Journal of Electrical and Computer Engineering Systems publishes original research in the form of full papers, case studies, reviews and surveys. It covers theory and application of electrical and computer engineering, synergy of computer systems and computational methods with electrical and electronic systems, as well as interdisciplinary research.

Topics of interest include, but are not limited to:

- Power systems
- Renewable electricity production
- Power electronics
- Electrical drives
- Industrial electronics
- Communication systems
- Advanced modulation techniques
- RFID devices and systems
- Signal and data processing
- Image processing
- Multimedia systems
- Microelectronics
- Instrumentation and measurement
- Control systems
- Robotics
- Modeling and simulation
- Modern computer architectures
- Computer networks
- Embedded systems
- High-performance computing
- Parallel and distributed computer systems
- Human-computer systems
- Intelligent systems
- Multi-agent and holonic systems
- Real-time systems
- Software engineering
- Internet and web applications and systems
- Applications of computer systems in engineering and related disciplines
- Mathematical models of engineering systems
- Engineering management
- Engineering education

Paper Submission

Authors are invited to submit original, unpublished research papers that are not being considered by another journal or any other publisher. Manuscripts must be submitted in doc, docx, rtf or pdf format, and limited to 30 one-column double-spaced pages. All figures and tables must be cited and placed in the body of the paper. Provide contact information of all authors and designate the corresponding author who should submit the manuscript to https://ijeces.ferit.hr. The corresponding author is responsible for ensuring that the article's publication has been approved by all coauthors and by the institutions of the authors if required. All enquiries concerning the publication of accepted papers should be sent to ijeces@ferit.hr.

The following information should be included in the submission:

- paper title;
- full name of each author;
- full institutional mailing addresses;
- e-mail addresses of each author;
- abstract (should be self-contained and not exceed 150 words). Introduction should have no subheadings;
- manuscript should contain one to five alphabetically ordered keywords;
- all abbreviations used in the manuscript should be explained by first appearance;
- all acknowledgments should be included at the end of the paper:
- authors are responsible for ensuring that the information in each reference is complete and accurate. All references must be numbered consecutively and citations of references in text should be identified using numbers in square brackets. All references should be cited within the text;
- each figure should be integrated in the text and cited in a consecutive order. Upon acceptance of the paper, each figure should be of high quality in one of the following formats: EPS, WMF, BMP and TIFF;
- corrected proofs must be returned to the publisher within 7 days of receipt.

Peer Review

All manuscripts are subject to peer review and must meet academic standards. Submissions will be first considered by an editor-in-chief and if not rejected right away, then they will be reviewed by anonymous reviewers. The submitting author will be asked to provide the names of 5 proposed reviewers including their e-mail addresses. The proposed reviewers should be in the research field of the manuscript. They should not be affiliated to the same institution of the manuscript author(s) and should not have had any collaboration with any of the authors during the last 3 years.

Author Benefits

The corresponding author will be provided with a .pdf file of the article or alternatively one hardcopy of the journal free of charge.

Units of Measurement

Units of measurement should be presented simply and concisely using System International (SI) units.

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003

Published: semiannually

Copyright

Authors of the International Journal of Electrical and Computer Engineering Systems must transfer copyright to the publisher in written form.

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.

Postal Address

Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b
31000 Osijek, Croatia

IJECES Copyright Transfer Form

(Please, read this carefully)

This form is intended for all accepted material submitted to the IJECES journal and must accompany any such material before publication.

TITLE OF ARTICLE (hereinafter referred to as "the Work"):

COMPLETE LIST OF AUTHORS:

The undersigned hereby assigns to the IJECES all rights under copyright that may exist in and to the above Work, and any revised or expanded works submitted to the IJECES by the undersigned based on the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the complete Work and all incorporated parts of the Work. Otherwise he/she warrants that necessary permissions have been obtained for those parts of works originating from other authors or publishers.

Authors retain all proprietary rights in any process or procedure described in the Work. Authors may reproduce or authorize others to reproduce the Work or derivative works for the author's personal use or for company use, provided that the source and the IJECES copyright notice are indicated, the copies are not used in any way that implies IJECES endorsement of a product or service of any author, and the copies themselves are not offered for sale. In the case of a Work performed under a special government contract or grant, the IJECES recognizes that the government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official government purposes only, if the contract/grant so requires. For all uses not covered previously, authors must ask for permission from the IJECES to reproduce or authorize the reproduction of the Work or material extracted from the Work. Although authors are permitted to re-use all or portions of the Work in other works, this excludes granting third-party requests for reprinting, republishing, or other types of re-use. The IJECES must handle all such third-party requests. The IJECES distributes its publication by various means and media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various collections, databases and other publications. The IJECES publisher requires that the consent of the first-named author be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes. If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IJECES publisher assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

Authors of IJECES journal articles and other material must ensure that their Work meets originality, authorship, author responsibilities and author misconduct requirements. It is the responsibility of the authors, not the IJECES publisher, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

- The undersigned represents that he/she has the authority to make and execute this assignment.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
- The undersigned agrees to indemnify and hold harmless the IJECES publisher from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

____________________________ ____________________________

Author/Authorized Agent **Date**

CONTACT
International Journal of Electrical and Computer Engineering Systems (IJECES)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek
Kneza Trpimira 2b
31000 Osijek, Croatia
Phone: +38531224600,
Fax: +38531224605,
e-mail: ijeces@ferit.hr