Wireless Communications and Mobile Computing

Wireless Networking Technologies for Smart Cities

Lead Guest Editor: Jaime Lloret
Guest Editors: Syed H. Ahmed, Danda B. Rawat, Waleed Ejaz, and Wei Yu

Wireless Networking Technologies for Smart Cities

Wireless Communications and Mobile Computing

Wireless Networking Technologies for Smart Cities

Lead Guest Editor: Jaime Lloret
Guest Editors: Syed H. Ahmed, Danda B. Rawat, Waleed Ejaz, and Wei Yu

Editorial Board

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 1865908, 3 pages
https://doi.org/10.1155/2018/1865908

Editorial

Editorial on Wireless Networking Technologies for Smart Cities

Jaime Lloret,[1] **Syed Hassan Ahmed**,[2] **Danda B. Rawat**,[3] **Waleed Ejaz**,[4] **and Wei Yu**[5]

[1]*Universitat Politecnica de Valencia, Valencia, Spain*
[2]*Georgia Southern University, USA*
[3]*Howard University, Washington, USA*
[4]*Ryerson University, Toronto, Canada*
[5]*Towson University, Maryland, USA*

Correspondence should be addressed to Jaime Lloret; jlloret@dcom.upv.es

Received 7 August 2018; Accepted 7 August 2018; Published 1 October 2018

Smart cities are becoming reality, due to the enormous research into the technology enablers and development of the Internet of Things (IoT) enabling a multitude of applications, which are built around various types of sensors. To manage the increased population of cities, there is a need to have more sustainable, environmentally-, and economically-friendly smarter cities and technologies. Moreover, the number of wireless technologies available for smart cities is increasing hugely [1]. Recently, despite the IoT devices constraints, such as their energy consumption [2], a number of smart devices appeared on a large scale, such as smart beds for monitoring and health data gathering during sleeping, like the SleepNumber smart bed, the Kolibree smart toothbrush, and the Belkin smart saucepan. These are few, with more applications and products to be showcased in the recent future.

The growing footprint of ultra-high-speed broadband networks, pervasive wireless networks, cloud computing, crowd sensing, and software-defined infrastructure connect smart/mobile devices to generate relevant city data on a massive scale. These advances will enable transformative applications and services that will enhance the quality of life while addressing important priorities such as real-time tracking, security, authenticity, and availability of classified information to the decision makers. Moreover, it will improve e-health services [3]. Similarly, to make a smart city, a strong communications infrastructure is required for connecting smart objects, people, and sensors. For instance, audio and video sensors support a variety of safety (monitoring) and other applications. Communication within cities involves multiple aggregations and access networks that can be either public or private. A city may gather data from smart devices and sensors embedded in the roadways, power grids, buildings, and other assets. It shares that data via a smart communications system that is typically a combination of wired and wireless networks, and in some cases mobile networks like Vehicular Ad Hoc Networks (VANETs) [4]. It then uses smart software to create valuable information and digitally enhanced services such as health care assistance, security and safety, real-time traffic monitoring, and managing the environment. The special issue contains 18 papers; each paper covers the subject from different prospective and thus offers readers a holistic view of different wireless technologies for smart cities.

Wireless sensor networks (WSNs) and IoT are key for the success of smart cities. In a paper entitled "Wireless Networking Performance in IoT Using Adaptive Contention Window", R. M. Bhavadharini et al. proposed an adaptive contention window to reduce the MAC overhead and retransmissions by determining active queue size at the contending nodes and the energy level of the nodes to improve TCP Performance. A novel indoor positioning system that considers the nonlinear discriminative feature extraction of received signal strength (RSS) using kernel local Fisher discriminant analysis (KLFDA) has been proposed by S. Imran and Y.-B. Ko in a paper entitled "A Novel Indoor

Positioning System Using Kernel Local Discriminant Analysis in Internet-of-Things". The discriminating information in the RSS is reorganized and maximally extracted using KLDFA. In an article by M. Xie et al. entitled "Weight-Aware Sensor Deployment in Wireless Sensor Networks for Smart Cities", a reliability model of sensing area is established based on the sensing accuracy of the sensor. The sensing area is divided into sensing grids and different weights are allocated for those grids. An optimization problem is also formulated to maximize the trust degree of the WSNs for the efficient deployment of sensors.

Cellular networks are of paramount importance for the smart cities. In an article by S. Tahira et al. entitled "Handover Based IMS Registration Scheme for Next Generation Mobile Networks", a lightweight fast IMS mobility registration scheme is proposed to avoid unnecessary conventional registration phases such as security associations, authentication, and authorization. In an article entitled "A Reinforcement Learning Approach to Access Management in Wireless Cellular Networks" J. Moon and Y. Lim proposed the use of access class barring (ACB) to restrict the number of access attempts allowed in data transmission by utilizing strategic parameters and modeled the problem of determining the strategic parameters with a reinforcement-learning algorithm.

Cognitive radio networks can address spectrum scarcity issues in smart cities. In an article by A. Kheaksong et al. entitled "Multicriteria Parent Selection Using Cognitive Radio for RPL in Smart Grid Network", a parent selecting mechanism is proposed to improve workload balancing and lifetime differences of all meters. The proposed mechanism is based on three criteria including expected transmission count, residual energy, and expected transmission time. In an article entitled "Adaptive and Blind Wideband Spectrum Sensing Scheme Using Singular Value Decomposition" Z. Hu et al. proposed an adaptive and blind reduced multiple measurement vectors boost (ABRMB) scheme based on singular value decomposition for wideband spectrum sensing.

Energy-efficient protocols and algorithms are also of prime importance for smart cities. In article entitled "EE-MRP: Energy-Efficient Multistage Routing Protocol for Wireless Sensor Networks" M. K. Khan et al. proposed an energy-efficient routing protocol for wireless sensor networks. The proposed protocol consists of a routing algorithm for the transmission of data, cluster head selection algorithm, and a scheme for the formation of clusters. An energy-efficient unequal chain length clustering (EEUCLC) protocol is proposed by M. Baniata and J. Hong in a paper entitled "Energy-Efficient Unequal Chain Length Clustering for Wireless Sensor Networks in Smart Cities". The protocol consists of a suboptimal multihop routing algorithm to reduce the burden on the cluster head and a probability-based cluster head selection algorithm to prolong the network lifetime.

Smart cities also have to deal with a Big Data for performance enhancement and decision-making. A smart city framework based on Big Data analytics is proposed by B. N. Silva et al. in an article entitled "Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making". The proposed framework operates on (i) data generation and acquisition, (ii) data management and processing, and (iii) application level initiates execution of the events. M. M. Iqbal et al. in an article entitled "Augmenting High-Performance Mobile Cloud Computations for Big Data in AMBER" proposed an innovative execution of assisted model building with energy refinement (AMBER) with force field empirical formula, using message passing interface (MPI) infrastructure on high-performance mobile cloud computing. The main objective of this work is facilitating research community with a mobile cloud of laptops capable of doing the heavy processing.

Security and privacy in smart cities have also been addressed in recent years. In an article entitled "TrustyFeer: A Subjective Logic Trust Model for Smart City Peer-to-Peer Federated Clouds" by H. Kurdi et al. proposed a trust management system called TrustyFeer that allows peers to evaluate the trustworthiness of other peers based on subjective logic opinions, formulated using peers' reputations and service level agreements (SLAs). A framework for performing a novel, template-based fuzzing technique on the message queue telemetry transport is proposed by S. H. Ramos et al. in an article entitled "MQTT Security: A Novel Fuzzing Approach". An agent-based open-source framework for testing different policies for normalizing and controlling self-reported priorities, with its simulator called ABS-smartPriority is proposed by I. García-Magariño and R.l Lacuesta in an article entitled "ABS-SmartPriority: An Agent-Based Simulator of Strategies for Managing Self-Reported Priorities in Smart Cities". This simulator is illustrated by two different strategies, i.e., smart policy for tracking this history of citizens and control mechanism for the priorities self-reported by the users. In order to reduce computation cost and achieve better security, in "Lightweight Data Aggregation Scheme against Internal Attackers in Smart Grid Using Elliptic Curve Cryptography" D. He et al. proposed a lightweight data aggregation scheme against internal attackers in the smart grid environment using elliptic curve cryptography.

The last group of articles in this special issue focus on wireless technologies for different applications in smart cities. An approach for mass-customization with the integration of smart retail and smart production is proposed by M. Karaköse and H. Yetiş in an article entitled "A Cyberphysical System Based Mass-Customization Approach with Integration of Industry 4.0 and Smart City". The objective is to reduce the waiting time of customers. A distributed group-based mobility management scheme is proposed by M. Gohar et al. in a paper entitled "Distributed Group-Based Mobility Management Scheme in Wireless Body Area Networks". A local mobility anchor (LMA) function is implemented by each mobile access gateway (MAG) and the handover operation is performed between two neighbouring MAGs without the help of LMA. In "Semantic Interoperability in Heterogeneous IoT Infrastructure for Healthcare" S. Jabbar et al. proposed an IoT based semantic interoperability model (IoT-SIM) to provide semantic interoperability among heterogeneous IoT devices in health care domain.

All of the papers in special issue highlight the importance of wireless technologies for the smart cities. In addition, authors investigated the issues related to security and privacy and Big Data analysis in smart cities. In a nutshell, the integration of new wireless technologies with the evolved current ones can bring tremendous improvement in future smart cities.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

We also would like to thank editorial staff, editor in chief, contributing authors, reviewers, and production staff for bringing our special issue to the level of success.

Jaime Lloret
Syed Hassan Ahmed
Danda B. Rawat
Waleed Ejaz
Wei Yu

References

[1] L. Garcia, J. M. Jiménez, M. Taha, and J. Lloret, "Wireless Technologies for IoT in Smart Cities," *Network Protocols and Algorithms*, vol. 10, no. 1, pp. 23–64, 2018.

[2] S. Rani, R. Talwar, J. Malhotra, S. H. Ahmed, M. Sarkar, and H. Song, "A novel scheme for an energy efficient internet of things based on wireless sensor networks," *Sensors*, vol. 15, no. 11, pp. 28603–28626, 2015.

[3] A. Rghioui, S. Sendra, J. Lloret, and A. Oumnad, "Internet of Things for Measuring Human Activities in Ambient Assisted Living and e-Health," *Network Protocols and Algorithms*, vol. 8, no. 3, pp. 15–28, 2016.

[4] Y. He, Z. Wei, G. Du, J. Li, N. Zhao, and H. Yin, "Securing cognitive radio vehicular Ad hoc networks with fog computing," *Ad Hoc & Sensor Wireless Networks*, vol. 40, no. 1-2, pp. 73–95, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 7248040, 9 pages
https://doi.org/10.1155/2018/7248040

Research Article

Wireless Networking Performance in IoT Using Adaptive Contention Window

R. M. Bhavadharini,[1,2,3] S. Karthik,[1,2,3] N. Karthikeyan,[1,2,3] and Anand Paul[1,2,3]

[1]*Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India*
[2]*SNS College of Engineering, Coimbatore, India*
[3]*The School of Computer Science and Engineering, Kyungpook National University, Republic of Korea*

Correspondence should be addressed to Anand Paul; paul.editor@gmail.com

Received 18 November 2017; Revised 22 March 2018; Accepted 8 April 2018; Published 3 July 2018

Academic Editor: Jaime Lloret

Internet of Things (IoT) network contains heterogeneous resource-constrained computing devices which has its unique reputation in IoT environments. In spite of its distinctiveness, the network performance deteriorates by the distributed contention of the nodes for the shared wireless medium in IoT. In IoT network, the Medium Access Control (MAC) layer contention impacts the level of congestion at the transport layer. Further, the increasing node contention at the MAC layer increases link layer frame drops resulting in timeouts at the transport layer segments and the performance of TCP degrades. In addition to that, the expiration of maximum retransmission attempts and the high contentions drive the MAC retransmissions and the associated overheads to reduce the link level throughput and the packet delivery ratio. In order to deal with aforementioned problems, the Adaptive Contention Window (ACW) is proposed, which aims to reduce the MAC overhead and retransmissions by determining active queue size at the contending nodes and the energy level of the nodes to improve TCP performance. Further, the MAC contention window is adjusted according to the node's active queue size and the residual energy and TCP congestion window is dynamically adjusted based on the MAC contention window. Hence, by adjusting the MAC Adaptive Contention Window, the proposed model effectively distributes the access to medium and assures improved network throughput. Finally, the simulation study implemented through ns-2 is compared with an existing methodology such as Cross-Layer Congestion Control and dynamic window adaptation (CC-BADWA); the proposed model enhances the network throughput with the minimal collisions.

1. Introduction

In the IoT realm, there are still many challenges open in innovative research as well as technical solutions to interoperability, integration, and interconnection of heterogeneous IoT systems, at a different level. Homogeneous cross platform domain in various IoT applications hinders the developer with various challenges including the cost, scalability, and reusability of IoT solutions. Figure 1 scheme contributes to the wireless technology which in turn supports the interoperability of IoT systems.

A smart object can be a sensor, a thing, or any physical device that has the capability to sense the environment, to collect the data from it, and to interact and communicate with any other physical device. These devices are called smart since they show intelligent behavior in connection and interaction with other devices via some wireless protocol and operate interactively and autonomously over the Internet [1–4]. Examples of these devices include, but are not limited to, smart watches, building automation sensors, medical and healthcare devices, smartphones, smart TVs, vehicles, and security system (Figure 2). These devices need to transmit the data by capturing the medium in a fair manner. The contention window in the MAC layer provides backoff time when the medium is busy. The backoff time is set according to node's queue size and the energy level to reduce MAC layer retransmissions. This current research approach in Adaptive Backoff Window for MAC model can contribute to the overall systems.

The distributed access to the shared wireless medium is controlled by Medium Access Control (MAC) protocol. Ad hoc based networks deploy IEEE 802.11 standard MAC

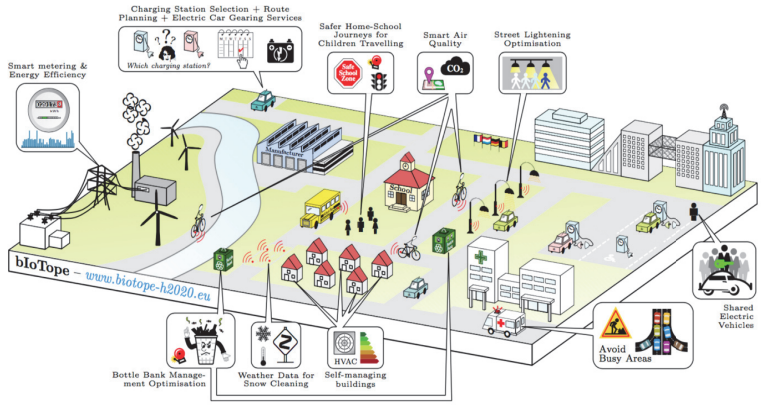


Figure 1: Pan European Interoperable IoT System.

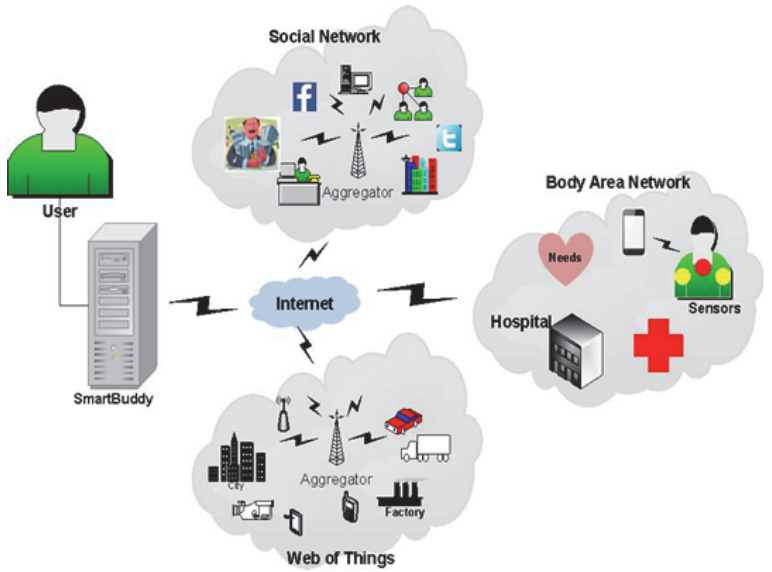


Figure 2: An overview of designed system.

protocol for ascertaining single hop reliability. IEEE 802.11 MAC uses two different access mechanisms. Distributed Coordinated Function (DCF) allows asynchronous data transfer from the nodes. No controlling infrastructure is used in DCF communication. Point Coordinated Function (PCF), on the other hand, has a central coordinator to control access to the shared medium.

The basic access mechanism used in DCF is Carrier Sense Multiple Access Collision Avoidance (CSMA/CA). In CSMA/CA channel state is sensed by the node before data transmission. The transmission is allowed or deferred based on either the channel is free or idle. Nodes in the sensing range of each other end up with collided packets when they sense the medium as idle and transmit at the same time. CSMA/CA uses a positive acknowledgment from the receiver to guarantee the reception of the transmitted frame. DCF features an exchange of control packets Ready to Send/Clear to Send (RTS/CTS) between the sender and the receiver to alleviate the hidden node or hidden terminal problem [5].

To transmit a frame, the sender transmits RTS signal to the receiver if the medium is sensed idle after DIFS interval time and random backoff time (BT). BT value is decremented after every slot if the medium is found to be idle; otherwise, BT is frozen until next DIFS (DCF Interframe Spacing) period of channel's idle time. The random value of backoff time reduces the chances of collisions among the frames. The receiver responds with CTS signal after SIFS (Short Interframe Spacing) period when the medium is free. After receiving CTS from the receiver, the sender sends DATA frame after SIFS and ACK are received from the receiver after SIFS. This sequence of message exchange is termed as four-way handshaking and successful data transmission includes DIFS+RTS+CTS+DATA+ACK+3SIFS time period.

The following sequence depicts successful frame transmission of IEEE 802.11 RTS/CTS access method as depicted in Figure 3:

t0-t1: STA1 senses the medium as idle for DIFS time

t1-t2: STA1 enters the contention window

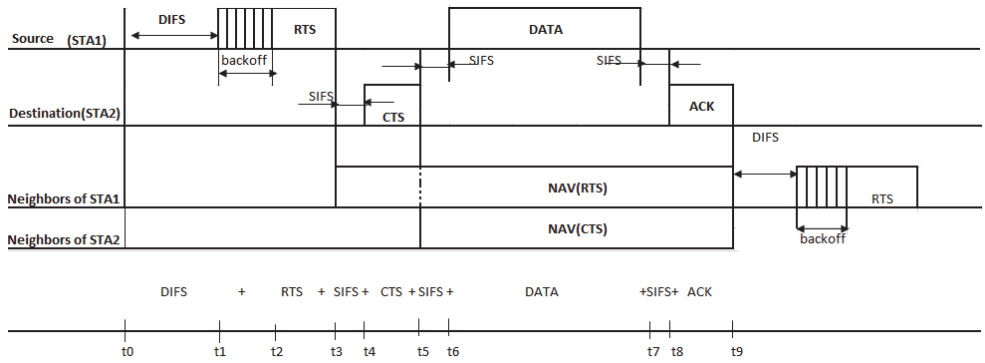


Figure 3: IEEE 802.11 DCF using RTS / CTS.

t2-t3: STA1 broadcast RTS

t3-t4: SIFS for transmitting CTS

t4-t5: CTS from STA2

t5-t6: SIFS before transmitting DATA

t6-t7: DATA transmitted from STA1 to STA2

t7-t8: SIFS before ACK

t8-t9: ACK sent from STA2 to STA1

If the RTS sender does not receive CTS for transmitted RTS or ACK for DATA frame, then the retransmission is triggered until the maximum retransmission limit (7 for control frames and 4 for data frames) after which the frame (either control frame or data frame) is dropped. This scenario is presumed as packet drop due to link failure at the MAC level and considered as the route failure at the network layer. As a result, TCP runs over IEEE 802.11 suffer a timeout because of the MAC contention losses and are wrongly perceived as congestion. Thus, congestion control measures are invoked resulting in retransmissions at the transport layer and degraded network throughput. Therefore, the contention in MANET impacts the TCP performance clearly. MAC layer uses Binary Exponential Backoff algorithm [BEB] that selects backoff counter value as *random [0, CW]* where CW is the current contention window. $CW \in (CW_{min}, CW_{max})$ and the initial value of CW is set to CW_{min} and doubled after every unsuccessful attempt until reaching a value of CW_{max} [6, 7]. BEB prevents simultaneous access to the wireless medium by randomizing the access times through random backoff procedure. The values of CWmin for 802.11a and 802.11b are 16 and 32, respectively. The value of CWmax is 1024 in both cases [8, 9].

BEB works as follows:

$$CW = \begin{cases} CW_{min} & \text{Initial value} \\ CW_{min} & \text{After successful frame transmission or max retry limit} \\ \text{Min}\left(2 * CW_{old}, CW_{max}\right) & \text{After unsuccessful frame transmission} \end{cases} \quad (1)$$

$BT = random[0, CW] * Channel\_slot\_time$.

Channel_slot_time is dependent on the physical layer characteristic of the node.

In BEB, the nodes with successful transmissions get the minimum backoff time, leaving them to access the medium with fewer contention slots. Other nodes deferring access to the shared medium and their increased contention window suffer more packet drops after maximum retry limit (short retry count for the control signals and long retry count for data signals). Retransmissions initiated for the dropped frames at the MAC level and unsuccessful retransmissions after retry limit result in a timeout at the TCP sender. As a result, TCP sender calls for false congestion control measure which worsens the condition. Added to this, MAC layer drops are also misinterpreted as a lost link and route rediscovery process is initiated which increases overhead in the network. Therefore, it is clear that CW value has an impact on congestion at the transport layer. Thus, if proper value setting of a contention window (CW) at the MAC layer is not suitably considered, the resulting network performance is seriously degraded.

The proposed MAC model ACW analyzes the current state of the network as well as the node to predict CW and backoff window values. Rather than fully depending on transport layer factors to decide congestion, ACW brings in cross-layer parameters for considering collision lead congestion. ACW sets the CW based on the node's queue size and the node's residual energy level. The MAC parameter CW is then used to adjust the transport layer congestion window CWND. Proper setting of CW reduces MAC collisions and retransmissions. Consequently, the packet delivery ratio increases and TCP throughput also gets improved.

2. Related Work

To improve the network performance, various research studies have been carried out on IEEE 802.11 for determining the proper value for CW. To resolve the problem of BEB, the effort taken in [10] described a contention window adjustment scheme called MILD (Multiplicative Increase and Linear Decrease) in which the CW is multiplied by 1.5 on collision and decreased by 1 on successful transmission. It includes a CW copy mechanism to overhear the copy of CW of successful transmission. It prevents quick escalation to minimum CW after successful transmission, thus improving the performance. Authors of [11] contributed Exponential Increase and Exponential Decrease (EIED) that improved the performance when more competing nodes exist to access medium in a wireless region. EIED doubled the contention window after the collision and halved after a successful transmission. An extension of MILD proposed in [12] is Linear/Multiplicative Increase and Linear Decrease (LMILD). The sender increases CW multiplicatively in account of a collision. Any node overhearing collision sets CW linearly and successful nodes decrease CW linearly.

In [13], the authors predicted the amount of outstanding data in the network with the help of contention delay experienced by the packets and the throughput achieved by the connections. Along the traveling path, every packet includes contention delay per hop and the receiving end estimates the average contention delay per hop. The variation in the successive RTT found by the receiver is then advertised to the sender to adjust the sending rate. They studied TCP intraflow instability in detail and the proposed method dynamically adjusts the TCP transmission rate to avoid excessive build-up packets in the queue. The research in [14] discusses the cross-layer contention control method to alleviate CWND overshooting problem. It measured the contention RTT through forwarding data frame and the backward ACK at the medium access layer and considered the variance in contention RTT per hop (VCRH) as the contention status.

Upon reception of new ACK, if VCRH exceeds the threshold value, congestion window is reduced by 1 MSS (Maximum Segment Size). If RTO expires, CWND is set to $2 * \mathrm{MSS}$. Along with other parameters, VCRH and timeout are quantified to congestion window adaptation.

Fu et al. [15] authored two mechanisms to improve the TCP to react to link overload: link RED (LRED) and adaptive pacing. Similar to random early detection for wired network, LRED drops frames based on a number of average MAC retransmission count. With increasing retry attempt, the packet drop probability increases because this situation indicates congestion. The mechanism adaptive pacing adds additional sender's MAC backoff time in addition to the current deferral time.

Saraireh et al. [16] employed an approach to improve IEEE 802.11 quality of service by fine-tuning the contention window dynamically. ToS (type of service) field of IP packet is used to classify high and low priority traffic. Collision Rate (CR) and Collision Rate Variation (CRV) are the parameters used for CW setting. High priority traffic is increased for smaller values of CR and CVR. In highly loaded network traffic, CW of the high priority traffic gets increased to reduce the delay. Differentiated services are prompted to access the medium with less collision rate both at light load and highly loaded network conditions. Knag et al. [17] examined estimation based backoff algorithm that estimates the number of idle slots during the backoff period. When there is a dramatic change in a number of active contending nodes or amount of traffic, the contention window is appropriately chosen.

Al-Hubaishi, Mohammed, et al. [18] suggested Enhanced Binary Exponential Backoff algorithm (E-BEB) that sets CW based on backoff time threshold value. The backoff threshold value is adaptively adjusted based on CW. It increases CW upon unsuccessful transmission and decreases to zero upon successful transmission. Authors [19] took into account the network history to optimize CW size. They implemented a channel state vector to store the network levels and based on the network level, CW is set.

Qin Yu et al. [20] projected the average idle interval time of an active node to dynamically adjust its contention window. They suggested this fair index to sense the network condition and less sensitive to the frame length and number of competing nodes. In [21], the authors proposed a dynamic control backoff time algorithm to handle nonsaturated traffic load of the network. They described the backoff algorithm with CW-Threshold as a reference value. When current CW is less than or equal to $\mathrm{CW}_{\mathrm{Th}}$, the CW is incremented exponentially for unsuccessful transmission and linearly decremented by 1 for successful transmission. When the network load goes high, then CW is incremented exponentially and added with two for transmission failure and linearly decremented by two. Gaurav et al. [22] presented a cross-layer mechanism for TCP performance improvement. They focused on link level contention, route failure, and congestion estimation to infer TCP sender about the network congestion.

Bhalerao et al. [23] designed and implemented CoCoA (Constrained Application Protocol Congestion Control Advanced) 4-state strong, an adaptation of 4 state estimator for variable backoff to improve the throughput. The proposed work has clearly distinguished between wireless losses and congestion losses. Rahman et al. [24] proposed a MAC strategy to mitigate the congestion issue in IoT enabled Intravehicular networks.

The research studies discussed above reveal that the value set for the contention window at the MAC layer of IEEE 802.11 has a significant improvement in the performance of wireless local area networks. To resolve the fairness issue among the competing nodes, the CW value is set based on the network condition. To avoid collisions and the packet drop rate, the contending stations must have contention slots such that all stations have fair access right to the medium. When collisions are reduced, the amount of legitimate traffic increases leading to less misprediction of a link failure at the network layer and the timeout at the TCP sender.

The proposed work emphasizes adjusting contention window size adaptive to the prevalent condition of the network and the node itself; thus, excessive waiting backoff slots are reduced.

3. The Objective of the Proposed Work

The shared and distributed access to the medium in the wireless network reasons the MAC layer collisions when the network is highly loaded. The existing backoff algorithms reduce/increase the contention window using decrement/increment factor when the medium is sensed to be busy. The MAC layer collision resulting from unfair CW triggers retransmissions and amplifies packet drops at MAC layer. More packet drops at lower layer thus induce a timeout for TCP segment and TCP sender initiates the congestion control algorithm unnecessarily. Hence, it is obvious that proper setting of CW value impacts TCP's improved performance.

4. Proposed Work Description

The proposed ACW considers the queue occupancy level and the residual energy at the node to adjust the contention window. CW is set smaller when queue occupancy level is more and node's energy is less to allow the node to access the medium immediately to avoid packet drop at MAC level. Backoff time estimation uses link scheduling among the neighboring nodes. The average transmission rate of node's neighbor's plays vital role in determining backoff time for transmitting and receiving packets in the contention region.

5. Active Queue Parameter (AQP)

The link layer queue stores packets to be transmitted to the next immediate hop node. When the queue becomes full, the packets get dropped. The packets stored in the queue that is reaching its full occupancy level should be transmitted quickly to alleviate packet drop. To provide significance to the queue with more packets, the CW considers the available queue length at the MAC and network layer of the node. Let AQS be the available queue size i.e., remaining queue capacity at the node, let TQS be the total queue size, and let $\text{AQPavg}_{(n-1)}$ be the AQP till (n-1) transmissions and $a1$, $a2$ are the random values between [0, 1]; then

$$\text{Active Queue Param (AQP)} = \frac{\text{AQS}}{\text{TQS}} \tag{2}$$

The AQP uses exponential weighted moving average method to estimate the value.

$$AQP\,(n+1) = a1 * \left(\frac{AQS}{TQS}\right) + a2 * \left(AQPavg_{(n-1)}\right) \tag{3}$$

For every transmission, AQP is estimated at the transmitting node. The estimated current buffer availability indicates whether queue reaches its full capacity soon or it has sufficient room for more packets. The value of AQP is reaching 1, meaning that the queue becomes full; the packets should be allowed to contend for the medium with a minimum value for CW. Otherwise, the packets will be dropped triggering retransmissions and timeouts.

If(energy > $\text{Energy}_{\text{TH}}$)
{
$\text{NEP} = \text{w1} * 1 + \text{w2} * \left(\frac{Converted\_energy\_wrt\_OE}{OE}\right)$
}
Else(energy <= Min_Threshold)
{
$\text{NEP} = \text{w1} * \left(\frac{AE}{Min\_Threshold}\right)$
}

Box 1

Table 1: NEP values.

Available Energy (Joules)	NEP
95	0.98
50	0.8
15	0.51

6. Node Energy Parameter (NEP)

The energy level of the node is taken into account for adjusting the contention window of the node. When the node's energy level is getting reduced, the node is given priority to transmit the packet. A threshold energy value is used to estimate the NEP. When the node's energy level reduces beyond the threshold value, it is given high priority to access the medium for data transmission with CW set to minimum value. Let OE be the optimal energy, let IE be the initial energy at a node, and let AE be the available energy at a node; then NEP is estimated as in Box 1.

The Min_Threshold value considered for minimum contention window is 20 and the optimal energy (OE) considered for the optimal contention window is 80. NEP and Converted_Energy_wrt_OE are as follows:

$$NEP = w1\left(\frac{Current\_energy\_after\_Th}{Min\_Threshold}\right) + w2\left(\frac{Converted\_\text{Energy\_wrt\_OE}}{OE}\right) \tag{4}$$

$$Converted\_Energy\_wrt\_OE = \left(\frac{CE}{IE}\right) * (IE - Min\_Threshold) \tag{5}$$

The value obtained from ($Current\_energy\_after\_Th/Min\_Threshold$) defines the node's energy parameter when the energy level goes below the threshold energy value. The value obtained from ($Converted$_Energy_wrt_OE/OE) defines the node's energy parameter when the energy level is optimal.

Table 1 tabulates the NEP values for different residual energy levels at a node. Table 1 clearly reveals that NEP value is low for low energy at a node and thus the node is allowed

to access the medium with minimum CW value if it has the data packets to transmit.

Let α_1, α_2 be the weight factors. Now, the contention window is dynamically adjusted as follows:

$$\mathrm{CW}_{\mathrm{param}} = \alpha_1 \mathrm{NEF} + \alpha_2 \mathrm{AQP} \tag{6}$$

Let $\mathrm{CW}_{\mathrm{max}}$ be the maximum CW for the standard; then

$$\mathrm{CW} = \mathrm{CW}_{\mathrm{max}} * \mathrm{CW}_{\mathrm{param}} \tag{7}$$

The nodes with minimum energy and maximum buffer occupancy are allowed to contend for the medium quickly with a minimum value for CW. The minimum CW should allow the nodes to transmit the packets immediately by capturing the channel.

7. Slot Time

Total time is split into small time frames during which the nodes can sense for channel state and transmit the data. The proposed method implements link scheduling to assign links to the active transmissions. The total slot duration is shared among the neighbors intact and the node is given by

$$\text{Slot time} = \frac{\text{slot duration}}{\text{active transmissions}} \tag{8}$$

Active transmission refers to node's current data transmissions. To identify the slot time for a neighbor node, Average Packet Transmission Rate (APTR) over last unit time and the Average Packet Transmission Rate over a certain time are considered

$$\begin{aligned}\mathrm{APTR} = \beta 1.\left(\frac{No\_of\_Pack\_transmitted}{unit\_time}\right) \\ + \beta 2.\left(\frac{No\_of\_Pack\_transmitted}{Observation\ \_time}\right)\end{aligned} \tag{9}$$

Both ($No\_of\_Pack\_transmitted/unit\_time$) and ($No\_of\_Pack\_transmitted/Observation\ \_time$) denote rate of data transmission of the nodes. For the neighbors (n1,n2,. . ...ni) of node node n, the APTR is calculated as APTR1, APTR2, APTR3. . ..APTRi. Depending on the APTR value, the node's slot time is estimated. Let sum_APTR be the total of APTR of all active nodes, let $\mathrm{Neighbor\_share}_n$ be the transmission rate of node_n with respect to all neighbors, and let $\mathrm{Slottime}_n$ be the slot time for node_n, to transmit the data; then backoff time for node_n is given by

$$\begin{aligned}\mathrm{sum\_APTR} &= \mathrm{APTR}_1 + \mathrm{APTR}_2 + \mathrm{APTR}_3 + \ldots \\ &\quad + \mathrm{APTR}_n \\ \mathrm{Neighbor\_share}_n &= \frac{\mathrm{APTR}_i}{\mathrm{sum\_APTR}} \\ \mathrm{Slottime}_n &= \mathrm{Neighbor\_share}_n * \mathrm{slot\_duration} \\ \mathrm{BT}_n &= \mathrm{CW} * \mathrm{slottime}\end{aligned} \tag{10}$$

BT_n denotes the time the node has to wait for capturing the medium for transmission. The nodes with more data transmission rate will get a fairer share of the wireless medium than nodes with less data transmission.

8. TCP Congestion Window

To avoid excess data transmission rate into the network, the TCP CWND is adjusted to determine the number of packets that could be in the network. Instead of choosing a random value for MAC backoff window, the current network parameter based contention window setting reduces the collisions and thus the MAC retransmissions. The backoff time decides the contention severity level of the node and accesses the medium which in turn is used to regulate the TCP transmission rate into the network. Thus, TCP congestion window is set based on the cross-layer MAC parameter as

$$\mathrm{CWND} = \mathrm{CWND}_{\mathrm{max}} * \mathrm{CW}_{\mathrm{param}} \tag{11}$$

9. Performance Parameters

9.1. Throughput. It is defined as the amount of data packets received by the receiver in a unit of time. Generally it is measured in bits per second.

$$\text{Throughput} = \frac{\sum \mathrm{Number\_of\_Packets}_{\mathrm{received}}}{\sum \mathrm{Total\_time}} \tag{12}$$

9.2. Control Overhead. Control overhead is defined as the total number of control packets generated to transmit the data packets.

9.3. Normalized Overhead. Normalized overhead is the ratio of the amount of control packet sent to the amount of data packets received.

$$\begin{aligned}&\mathrm{Normalized\_Overhead} \\ &= \frac{\sum \mathrm{Number\_of\_Control\_Overhead}}{\sum \mathrm{Number\_of\_received\_Packets}}\end{aligned} \tag{13}$$

9.4. Average Energy Consumption. Average energy consumption is defined as the energy consumed by all the nodes in the network.

10. Results and Discussions

The proposed method ACW is an improvement of our previous work CC-BADWA. The results of ACW is compared with CC-BADWA and RACC (Receiver Assisted Congestion Control) congestion control methods. The results acquired show that ACW has bettered the results of other two existing methods. The analytical implementation was carried out using Network Simulator, 2. The proposed model has been analyzed in 1000 x 1000 topological area with 50 nodes moving with random waypoint model. The MAC standard IEEE 802.11 is used to access the medium.

Figure 4 illustrates that throughput of the proposed ACW outperforms other throughput models in the wireless environment. The proposed method evaluates CW over the network parameters and dynamically adjusts CWND not only based on segment loss but considering the contention

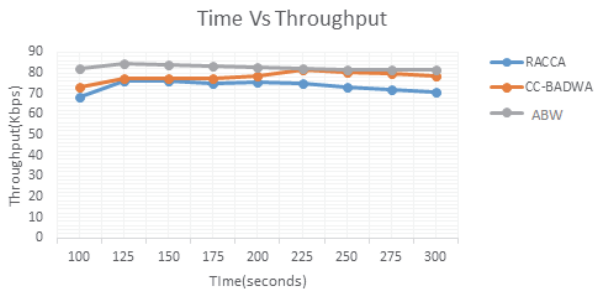


Figure 4: Comparison of throughput: RACC, CC-BADWA, and ACW.

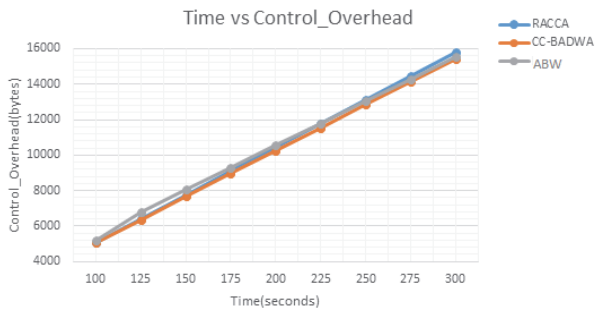


Figure 5: Comparison of control overhead: RACC, CC-BADWA, and ACW.

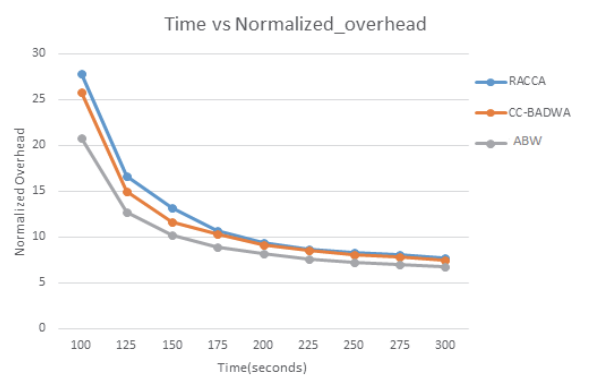


Figure 6: Comparison of normalized overhead: RACC, CC-BADWA, and ACW.

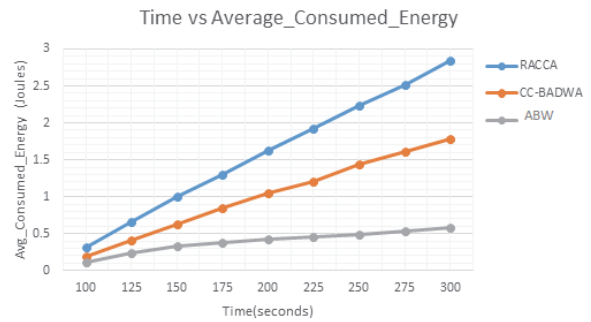


Figure 7: Comparison of Average_Consumed_Energy: RACC, CC-BADWA, and ACW.

status in the network, thus gaining improvement in TCP throughput. The collision probability is reduced because of the dynamic setting of CW value and avoiding false trigger of TCP congestion algorithm resulting in less number of retransmissions, whereas in conventional TCP, the contention losses are treated as congestion when there is no congestion. ACW uses optimal contention window size based on the network conditions leading to a fair share of the channel access to all contending nodes. The value of CWND depends on the value of CW and hence the outstanding packets in the network are also controlled by the TCP sender. Eventually, the sending rate adaptively increases or decreases based on contention level in the network. So the PDR and throughput in ACW are improved.

Figures 5 and 6 compare the control overhead and normalized overhead of the compared mechanism. The control overheads are almost the same in all three methods, whereas the normalized overhead is reduced in ACW because the rate of collisions is reduced and hence the packet drop is reduced. The backoff time is chosen depending on the neighbor node's transmission rate and the slot time. The node with more data will get longer slot time than the rest of the nodes. This adaptive backoff time reduces the random backoff time used in BEB and thus the number of RTS/CTS control messages is reduced. So the control packets exchanged for the retransmitted packets also reduced leading to less normalized overhead than the other two methods. Ultimately, the successful transmission rate increases reducing the retransmissions.

As shown in Figure 7, ACW reduces the average energy consumption since the dynamic setting of the backoff time is adaptive to current network condition. The Node Energy Parameter (NEP) estimated for contention window setting chooses the threshold value such that the node with low energy gets earlier access to the medium, thus avoiding packet losses and redundant transmissions. Also, the adaptive CW sets the value in line with network parameters not like BEB where CW is set to CWmin and $2 * CW$ after successful and unsuccessful transmissions, respectively. The conventional BEB causes starvation among the nodes that have lost the race and double CW every time. After the maximum retry count, the packet is dropped, timeouts at TCP trigger congestion control, and more energy consumption results. ACW's contention parameter impacts congestion window which interprets congestion accordingly.

11. Conclusion

The proposed Adaptive Backoff Window (ACW) for MAC model is to reduce MAC layer collisions by adaptively setting the value for a backoff window for every successful or unsuccessful MAC frame transmissions. The nodes willing to transmit the frame after waiting for a random backoff time

attempt to capture the medium to transmit the frames. When the medium is sensed busy, the random contention window size is chosen by the current network and node parameters. CW in ACW is set upon node's energy parameter, node's queue parameter, the active nodes in the region, and RP value. The dynamic value of backoff and CW decreases the MAC layer contentions and collisions which could otherwise have caused packet drops at the link layer and timeouts at the transport layer. Thus, the timeouts and unnecessary invocations of TCP congestion control algorithms are reduced. Thus, reductions in MAC retransmissions due to dynamic backoff improve the TCP performance. The extensive simulations carried out prove that ACW has reduced the packet drop ratio and end-to-end delay. The results clearly indicate that the number of successful transmissions is increased compared to the existing methods. Eventually, the MAC packet drops and timeouts are reduced. Since the adaptive CW indicates the present network condition, it is used to set the TCP CWND which decides the amount of data packets that could be in flight. Both CW and CWND decide the appropriate number of packets that can be in the network pipe so that the drop ratio and timeouts are reduced and TCP performance is improved in terms of network throughput.

Additional Points

Scope for Future Enhancement. The proposed work considers the MAC layer contention status to predict the data sending rate at the sender node. Future work will further proceed with the feedback information from the physical layer since the PHY layer characteristics can promptly reveal the wireless medium status to predict real congestion in the network.

Conflicts of Interest

The authors declare that the grant, scholarship, and/or funding mentioned in Acknowledgments does not lead to any conflicts of interest. Additionally, the authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (NRF-2017R1C1B5017464).

References

[1] A. Paul, A. Ahmad, M. M. Rathore, and S. Jabbar, "Smartbuddy: defining human behaviors using big data analytics in social internet of things," *IEEE Wireless Communications Magazine*, vol. 23, no. 5, pp. 68–74, 2016.

[2] A. Paul, T. Victoire, and A. Jeyakumar, "Particle swarm approach for retiming in VLSI," in *Proceedings of the 46th International Midwest Symposium on Circuits and Systems*, pp. 27–30, Cairo, Egypt, 2003.

[3] A. Paul, "Real-time power management for embedded M2M using intelligent learning methods," *ACM Transactions on Embedded Computing Systems*, pp. 1–22, 2014.

[4] A. Paul, "Graph based M2M optimization in an IoT environment," in *Proceedings of the Research in Adaptive and Convergent Systems (RACS '13)*, 2013.

[5] W. Liu, H. Jin, X. Wang, and M. Guizani, "A novel IEEE 802.11-based MAC protocol supporting cooperative communications," *International Journal of Communication Systems*, vol. 24, no. 11, pp. 1480–1495, 2011.

[6] K. Hong, S. K. Lee, K. Kim, and Y. H. Kim, "Channel condition based contention window adaptation in IEEE 802.11 WLANs," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 469–478, 2012.

[7] M. Barry, A. Campbell, and A. Veres, "Distributed control algorithms for service differentiation in wireless packet networks," in *Proceedings of the INFOCOM*, Anchorage, AK, USA, 2001.

[8] Y.-H. Zhu, X.-Z. Tian, and J. Zheng, "Performance analysis of the binary exponential backoff algorithm for IEEE 802.11 based mobile ad hoc networks," in *Proceedings of the IEEE International Conference on Communications (ICC)*, 2011.

[9] Q. Pang, S. C. Liew, J. Y. B. Lee, and V. C. M. Leung, "Performance evaluation of an adaptive backoff scheme for WLAN," *Wireless Communications and Mobile Computing*, vol. 4, no. 8, pp. 867–879, 2004.

[10] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: a Media Access Protocol for Wireless LANs," in *Proceedings of the SIGCOMM'94*, 1994.

[11] N. Song, B. Kwak, J. Song, and ME. Miller, "Enhancement of IEEE 802.11 distributed coordination function with exponential increase exponential decrease backoff algorithm," in *Proceedings of the 57th IEEE Semiannual Spring VTC*, vol. 4, pp. 2775–2778, 2003.

[12] A. Paul and S. Rho, "Probabilistic model for M2M in IoT networking and communication," *Telecommunication Systems*, vol. 62, no. 1, pp. 59–66, 2015.

[13] E. Hamadani and V. Rakocevic, "TCP contention control: A cross-layer approach to improve TCP performance in multihop ad hoc networks," in *Proceedings of the 5th Int. Conf. Wired/Wireless Internet Commun.*, pp. 1–16, 2007.

[14] X. M. Zhang, W. B. Zhu, N. N. Li, and D. K. Sung, "TCP congestion window adaptation through contention detection in Ad Hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4578–4588, 2010.

[15] Z. Fu, H. Luo, P. Zerfos, S. Lu, L. Zhang, and M. Gerla, "The impact of multihop wireless channel on TCP performance," *IEEE Transactions on Mobile Computing*, vol. 4, no. 2, pp. 209–221, 2005.

[16] M. Saraireh, J. AL-Saraire, and S. Saraireh, "A Novel Adaptive Contention Window Scheme for IEEE 802.11 MAC Protocol," *Trends in Applied Sciences Research*, vol. 9, no. 6, pp. 275–289, 2014.

[17] S.-W. Kang, J.-R. Cha, and J.-H. Kim, "A novel estimation-based backoff algorithm in the IEEE 802.11 based wireless network," in *Proceedings of the Consumer Communications and Networking Conference (CCNC)*, 2010.

[18] M. Al-Hubaishi, T. Alahdal, R. Alsaqour, A. Berqia, M. Abdelhaq, and O. Alsaqour, "Enhanced binary exponential backoff algorithm for fair channel access in the IEEE 802.11 medium access control protocol," *International Journal of Communication Systems*, vol. 27, no. 12, pp. 4166–4184, 2014.

[19] A. Balador, A. Movaghar, S. Jabbehdari, and D. Kanellopoulos, "A novel contention window control scheme for IEEE 802.11 WLANs," *IETE Technical Review*, vol. 29, no. 3, pp. 202–212, 2012.

[20] Q. Yu, Y. Zhuang, and L. Ma, "Dynamic contention window adjustment scheme for improving throughput and fairness in IEEE 802.11 wireless LANs," in *Proceedings of the Global Communications Conference (GLOBECOM)*, 2012.

[21] H. Alkadeki, X. Wang, and M. Odetayo, "Improving Performance of IEEE 802.11 By a Dynamic Control Backoff Algorithm Under Unsaturated Traffic Loads," *International Journal of Wireless & Mobile Networks (IJWMN)*, vol. 7, no. 6, 2015.

[22] G. Bhatia and V. Kumar, "CTCP: A Cross-Layer Information Based TCP for MANET," *International Journal of Ad Hoc, Sensor & Ubiquitous Computing*, vol. 5, no. 1, 2014.

[23] R. Bhalerao, S. S. Subramanian, and J. Pasquale, "An analysis and improvement of congestion control in the CoAP Internet-of-Things protocol," in *Proceedings of the Consumer Communications & Networking Conference (CCNC), 13th IEEE Annual. IEEE*, 2016.

[24] M. A. Rahman, M. N. Kabir, S. Azad, and J. Ali, "On mitigating hop-to-hop congestion problem in IoT enabled Intra-Vehicular communication," in *Proceedings of the Software Engineering and Computer Systems (ICSECS), 4th International Conference on. IEEE*, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5913836, 15 pages
https://doi.org/10.1155/2018/5913836

Research Article

Weight-Aware Sensor Deployment in Wireless Sensor Networks for Smart Cities

Mingshan Xie,[1,2] **Yong Bai**,[1] **Zhuhua Hu**,[1] **and Chong Shen**[1]

[1]*State Key Laboratory of Marine Resource Utilization in South China Sea, College of Information Science & Technology, Hainan University, Haikou 570228, China*
[2]*College of Network, Haikou University of Economics, Haikou 571127, China*

Correspondence should be addressed to Yong Bai; bai@hainu.edu.cn

Received 1 September 2017; Revised 16 January 2018; Accepted 19 February 2018; Published 4 April 2018

Academic Editor: Danda B. Rawat

During the construction of wireless sensor networks (WSNs) for smart cities, a preliminary survey of the relative criticalness within the monitored area can be performed. It is a challenge for deterministic sensor deployment to balance the tradeoff of sensing reliability and cost. In this paper, based on the sensing accuracy of the sensor, we establish a reliability model of the sensing area which is divided into sensing grids, and different weights are allocated to those grids. We employ a practical evaluation criterion using seesaw mapping for determining the weights of sensing grids. We further formulate and solve an optimization problem for maximizing the trust degree of the WSNs. With our proposed method, the efficient deployment of sensors can be realized. Simulation results show that our proposed deployment strategy can achieve higher trust degree with reduced sensor deployment cost and lower number of sensors at a certain miss probability threshold.

1. Introduction

There are many wireless sensor networks (WSNs) applications in smart cities, such as residential security, factory pollution monitoring, elevator monitoring, lamp control, power grid monitoring, fire monitoring, medical care, building monitoring, and air monitoring. During the construction of wireless sensor networks for smart cities, we can survey the terrain situations in advance. In the monitored region, data in some areas is more critical. For instance, in traffic accident monitoring, some sections, such as curves and ramps, are accident-prone areas. Since the accuracy of the sensors cannot be the same in each monitored area, the data collection capability of the critical areas should be considered in the deployment of sensors in these regions.

At present, there are many research works about sensor deployment. They focus on various aspects. Some studies are about routing of senses with the consideration of the energy problem [1]. The authors have optimized the sensor placement for indoor localization [2]. In the process of sensor deployment, the weight of the sensing region should be considered. In [3], Zou and Chakrabarty considered the most vulnerable regions in the sensor field when designing algorithms. In [4], the authors think it is necessary to locate the sensor nodes in the field of interest which refers to the preferred field of coverage. It is of practical significance that the sensor deployment should consider the criticalness of each subregion of the sensing area. It is not necessary to deploy sensors in every corner. We need to weigh the criticalness of each subregion in the monitored region in order to ensure the reliability of the data collected.

In our work, we present an optimization framework for sensor deployment. The proposed approach aims to improve the comprehensive performance of the sensor network and ensure the effective collection of data.

The contributions of this paper are listed as follows.

(1) In order to collect data effectively and reliably, considering the actual engineering requirements, this paper proposes a weight-aware sensor deployment method based on the weight of each sensing grid in the monitored region, which is the meshed region according to the sensor's sensing region.

(2) In this paper, the weight of the sensing grid is estimated from three aspects: the probability of abnormal events,

the consequence of abnormal events, and the tolerance time of abnormal events. A seesaw mapping is proposed to calculate the weights by taking into account the combined effect of the influencing factors, the cost of damage resulting from an abnormal event, and the tolerance time of an abnormal event.

(3) In this paper, the sensing accuracy and coverage degree are synthesized and the sensing trust degree model of the whole WSNs is proposed by taking the weight of sensing grid into account.

(4) Considering the fact that sensing areas are often subject to obstacles and sensing accuracy is limited, the corresponding algorithms are proposed, which makes the sensing trust degree calculation model more widely useful.

The remainder of this paper is organized as follows. In Section 2, we review prior research on the WSNs coverage and deployment. The model for the optimized sensor deployment is proposed in Section 3, and the estimation for the weight is in Section 4. The computing for the sensing trust degree of whole WSNs is also given in detail in Section 5. In Section 6, we describe the weight-aware sensor deployment algorithms in detail. In Section 7, we present the simulation results to evaluate the network performance for the proposed algorithms. Finally, we summarize the main results and give concluding remarks in Section 8.

2. Related Work

When sensor nodes are deployed in the monitored area, it is critical to calculate the weight of the sensing region. We divide the sensing region into many grids. An irregular sensor field is modeled as a collection of grids. The sensing matrix is formed by the weights of the sensing grids. Much research has considered dividing the sensing region into many grids. Xiao and Zhang proposed a divide-and-conquer-based surveillance framework, in which a large surveillance area is divided into small areas by critical points and critical lines [5]. In [6], the authors consider an optimization problem of how to partition the given field into multiple parcels and to deploy sensor nodes in some selected parcels such that the field information covered by the deployed sensor nodes meets the requirement. In [7], the performance analysis of different grid types (e.g., triangle, square, pentagon, hexagon, heptagon, and octagon) is discussed by computing their coverage area and efficient coverage area ratios.

In the smart cities, sensor deployment must take into account the nature of the terrain, for example, obstacles such as buildings and trees in the line of vision for IR sensors, uneven surfaces and elevations for hilly terrains, redundancy due to the likelihood of sensor failures, and the power needed to transmit information between deployed sensors and between a deployed sensor and the cluster head [8].

Sensor deployment methods consider obstacles. In [9], the authors develop a series of mechanisms to handle the obstacles in the 3D environment and propose a greedy heuristic. Chang et al. [10] presented the efficient obstacle-resistant robot deployment (ORRD) algorithm, which involves the design of a node localization policy, a serpentine movement policy, obstacle-handling rules, and boundary rules.

To ensure the efficient collection of data, the minimum reliability of wireless sensor networks needs to be considered. Long et al. [11] present an efficient data gathering scheme that guarantees the quality of service and optimizes the following network performance metrics as well as the end-to-end reliability in WSNs. In [12], Deif and Gadallah proposed an ant colony optimization (ACO) algorithm to solve the minimum cost reliability constrained sensor node deployment problem.

In summary, there has been a lot of work on the sensor deployment of the WSNs. Some papers are to solve the coverage problem [13–16]. However, most of the researches did not discuss the weight of the sensing field in detail. In fact, the weight of the sensing field has a great influence on the reliability and accuracy of data sampling. The relationship between factors which are involved in the estimation of weights is complicated. In order to solve this problem, we develop a novel seesaw mapping. The major issue of deploying sensors is equivalent to the problem of finding the location combination of the sensors. Considering the weights, we proposed a series of weight-aware sensor deployment algorithms in three cases which are the given total number of sensors, minimum sensing reliability, and some obstacles in the sensing region.

3. Deployment Model

The deployment of the sensor network should ensure that all the monitored regions are covered. The number of sensors and the cost of the network are also considered. We can estimate the covered area and the position of the sensors using the method based on the grids. The monitored region is divided into several grids in which the sensors are deployed [17].

The probability of abnormal events is very high in certain fields of the monitored region. In monitoring, it is necessary to obtain data from these areas more precisely. In order to obtain accurate monitoring data of an interesting area, several sensors are usually deployed. Sensor deployment requires accurate and computationally feasible sensor detection models. When the sensor is deployed in the critical area, the minimum sensing radius of the sensor r_s that does not obstruct the monitoring is considered as the effective sensing threshold of the sensor node of the whole network. The monitored area is divided into $M \times N$ sensing grids. We assume that each sensing grid is a regular inscribed quadrilateral of a circle whose diameter is $2r_s$. The index of the sensing grid is (i, j), where $1 \le i \le M$, $1 \le j \le N$. The model is illustrated in Figure 1. The sensing data matrix of the monitored area is set as $X \in \mathbb{R}^{M \times N}$.

In this work, we determine a sensing grid to deploy only one sensor. If an area needs more sensors, we can take measures of selecting smaller r_s or increase the sensor layout density.

Here, we adopt a mask operator $\Gamma(\cdot)$ to represent the procedure of deploying the sensors:

$$\Gamma\left(X \otimes W'\right) = X \otimes W' \otimes Q, \tag{1}$$

where W' represents the matrix of the sensing grids' weights and $\otimes$ represents the elementwise product of two matrices. Q is an $M \times N$ matrix defined by the following form:

$$Q(i,j) = \begin{cases} 1 & \text{A sensor has been deployed in the grid which is the } (i,j) \text{ th position of the } X \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where (i, j) denotes the index of the sensing grid in Q. The sensors localization set P is obtained by combining all the coordinates of the sensing grid that satisfy $Q(i, j) = 1$:

$$P = \{(i,j) \mid Q(i,j) = 1\}. \quad (3)$$

Now, we can use k to represent the total number of the elements of P. It is obvious that k is the total number of sensors.

The index of elements of Q is numbered by line priority. The index of the numbers is from 1 to $M \times N$. The essence of weight-aware sensor deterministic deployment optimization is a combinatorial optimization issue for the number of sensing grids in which the sensors are located. The combination scheme for the number of sensor positions with the best monitoring effect is selected from many combinations. The set of various combinations of sensors is

$$L = \left\{L_g \mid 0 \le g \le C_{M\times N}^{g},\ g \in Z\right\}, \quad (4)$$

where L_g stands for the gth combination of k numbers which is from 1 to $M \times N$. g is the index of various combinations.

$$F\left(L_g\right) = P, \quad (5)$$

where the function $F()$ is to convert a string of numbers into a series of coordinates in the sensing matrix X. Equation (5) denotes that the sensor localization set P is obtained by taking the gth combination scheme L_g.

For instance, the index set of the elements of Q is $\{(1,1),(1,2),(2,1),(2,2)\}$. The numbered set corresponding to Q is $\{1,2,3,4\}$. Now, we assume that $k = 3$ sensors that need to be deployed. We take any three numbers from the numbered set $\{1,2,3,4\}$ to form a combination. $C_4^3 = 4$. There are four cases of taking any three numbers from the numbered set. $L = \{L_1, L_2, L_3, L_4\} = \{\langle 1,2,3\rangle, \langle 1,2,4\rangle, \langle 2,3,4\rangle, \langle 1,3,4\rangle\}$. The second combination is used to explain the process. $g = 2$, and then $L_2 = \langle 1,2,4\rangle$. $F(L_2) = P = \{(1,1),(1,2),(2,2)\}$. The three sensors will be deployed to three grids whose indexes are $(1,1)$, $(1,2)$, and $(2,2)$.

There are many combinations of sensor localization. It becomes a critical issue for the weight-aware sensor deployment to choose the best one. What criterion is the best combination of the sensor localization? To solve this problem, we need to consider the weight of the sensing grid in the matrix X and ensure the effective and reliable data collection. We introduce the concept of the sensing trust degree of the whole WSN to evaluate deployment. It is related to the sensing reliability degree of the sensing grid, the sensing coverage rate of the whole WSN, and the weight of the sensing grid. In the following, we introduce these concepts.

Definition 1 (the sensing reliability degree of the sensing grid). This is defined as the probability that the data in the sensing grid can be effectively and successfully collected and transmitted to the sink node. The sensing reliability degree of the (i, j)th sensing grid is represented by $R(i, j)$.

$R(i, j)$ is the dual concept of data miss probability in sensing grid. The more the data points collected in the sensing region, the smaller the possibility of data loss and the greater the value of $R(i, j)$. $R(i, j)$ is the result of the combined action of all sensors that can sense the (i, j)th sensing grid. It is related to the total number of sensors that can cover the sensing grid and the sensors' sensing accuracies.

The more the sensor nodes that collect the same data are, the higher the reliability of the data acquisition $R(i, j)$ is. When one piece of data of the grid is collected by multiple sensors, if one of the sensors fails, it can be acquired by other sensors. For the critical areas of the monitored region, the reliability of data acquisition must be guaranteed.

Definition 2 (the sensing coverage degree of the sensing grid). The more powerful the sensor, the greater the area that it extends. The coverage ability of sensors is reflected by their sensing accuracy. The sensing coverage degree of the sensing grid is related to the sensing accuracy of sensors covered in this grid. The effective coverage degree produced by the most powerful sensor in this grid is used as the sensing coverage degree. The most powerful sensor's sensing accuracy is maximum. The sensing coverage degree of the (i, j)th sensing grid is represented by $CO(i, j)$. $CO(i, j) \times S(i, j)$ is defined as the effective coverage area of the (i, j)th sensing grid, where $S(i, j)$ denotes the area of the (i, j)th grid.

The sensing coverage rate of the whole WSNs is denoted by SCO.

$$SCO = \frac{COS}{AS}, \quad (6)$$

where AS denotes the area of the whole sensing region. COS denotes the effective coverage area of the whole WSN.

$$COS = \sum_{i=1}^{M}\sum_{j=1}^{N} CO(i,j) \times S(i,j). \quad (7)$$

The area where sensors overlap increases as the density of sensor deployment increases. As a result, the sensing coverage of the whole network is becoming lower. SCO reflects the density degree of sensor deployment.

Definition 3 (the sensing trust degree of the whole WSNs). It is the sum of all sensing reliabilities of sensing grids in a certain sensing coverage rate of the whole WSN. It is denoted by $TR(\cdot)$. The higher the value of $TR(\cdot)$ is, the higher the reliability of the data collected is, that is, the better the performance of the monitoring network is.

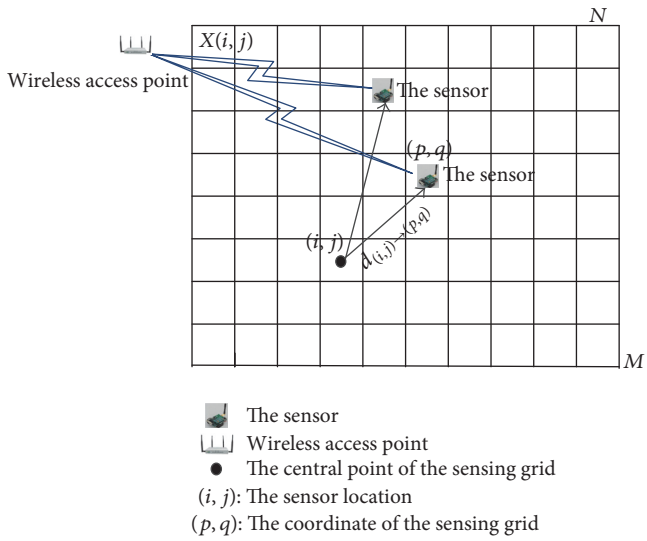


Figure 1: The sensing grids of the monitored region.

Finally, in the case that the number of sensors is given, we can get the best sensor deployment scheme by solving the following optimization problem:

$$P^* = \arg\max_P \quad TR\left(\Gamma\left(X \otimes W'\right)\right) \quad \text{s.t.} \quad k \ll M \times N. \tag{8}$$

In actual engineering, to ensure the effective collection of data from the monitored region, we also encounter the minimum requirement for grid sensing. It is discussed in the following.

In practice, the minimum sensor reliability $R_{\min}$ of sensing grid and the data loss probability threshold $M_{\min}$ in [8] are a pair of dual concepts. The larger the value of $R_{\min}$ is, the smaller the value of $M_{\min}$ is, and vice versa. In the case that the minimum sensor reliability is defined, the problem of weight-aware sensor deterministic deployment problem becomes

$$P^* = \arg\max_P \quad TR\left(\Gamma\left(X \otimes W'\right)\right) \quad \text{subject to} \quad R(i,j) \geq R_{\min}. \tag{9}$$

4. Estimation of the Weights

Weight estimation is a complex work which is investigated with many methods. In [18], Fujimoto et al. proposed a model reestimation method that involves the extraction of reliable characteristics using Gaussian pruning with weight normalization. In this paper, we hold the main factors that affect the weights. However, the relationship between the weight and the factors is not simply cumulative.

The weight of each sensing grid is affected by many factors. In this paper, we have found the three main factors: PL, the probability of an abnormal event occurrence; D, the cost of damage resulting from an abnormal event, that is, the extent of the impact; and T, the tolerance time of an abnormal event, that is, the period time over which the monitored object exhibits an abnormal event to cause great losses. These three quantities of monitoring objects can be estimated by fuzzy comprehensive evaluation method or by Delphi method.

Among them, the relationship between the cost of damage resulting from an abnormal event and the tolerance time of an abnormal event is very special. The two variables are independent and unrelated to each other, and they have the opposite direction of development. The problem of calculating their effects on computation of the weight is getting more and more attention. Since the two variables are independent and unrelated, their coordinates are perpendicular to each other. It is not appropriate to calculate the common effects of two independent variables in a summation method. In this paper, the characteristics of two independent and unrelated variables are analyzed, and a novel seesaw mapping is introduced on the basis of the seesaw model in [19, 20].

4.1. The Seesaw Mapping. In some application scenarios, two independent variables or entities affect another variable or entity together, but the direction of development of these two entities is opposite, and two variables have opposite orientations relative to the third variable or entity. It is necessary to determine the extent of the combined effect.

The condition of seesaw model in [19] is as follows: (1) two independent variables or entities are generally independent of each other; (2) two independent variables or entities affect another variable or entity together. For the two variables or entities that satisfy the condition of the seesaw model, a seesaw mapping can be applied. The seesaw model condition ensures that the two variables or entities can be mapped to Euclidean space. The values of the two coordinates in the Euclidean space reflect the values of the two variables at some certain state points and the values of two individual developments. It is necessary to map the common effects of two quantities in the Euclidean space to a single coordinate. We compare their values by the front and back relationships in a coordinate. This comparison is intuitive and clear. This series of processes is called the seesaw mapping.

We assume the variables A and B satisfied the condition of seesaw model. $A = \{A_i \in \mathbb{R} \mid i \in \mathrm{Z}\}$ *and* $B = \{B_i \in \mathbb{R} \mid i \in \mathrm{Z}\}$. The set of the seesaw pairs of A and B is set as $M = \{(A_i, B_i) \mid i \in T\}$, where (A_i, B_i) is a seesaw pair. One A_i corresponds to one B_i; namely, there is a one-to-one correspondence between A_i and B_i. T is the distribution sequence obeyed by A and B. $T \subset \mathbb{R}$. The distribution sequence can be either spatial or temporal or any other distribution. For example, we can use the seesaw mapping to study the changes of A and B over time. We can also use it to study the difference between A and B when they are at different points in space. i is the index of (A, B) in T. T is commonly $\{0, 1, 2, \ldots\}$ or $[c, d]$, where c can take 0 or $-\infty$ and d can be $+\infty$. If T is a countable set, we call T a discrete sequence. A_i denotes the value of variable A at point i in T. B_i denotes the value of variable B at point i in T.

Definition 4 (the state of the combined effect). This is the quantitative representation of the combined effect of A and B at the i point in T. It is represented by $G_i(A_i, B_i)$, and $G_i(A_i, B_i) \in \mathbb{R}$.

Definition 5 (the state space). This is the set of $G_i(A_i, B_i)$. One calls the state space of the combined effect of A and B. It is denoted by $G_T(A, B)$. $G_T(A, B) \subset \mathbb{R}$.

The seesaw mapping is represented as a mapping:

$$(A_i, B_i) \xrightarrow{f(*,*)} G_i(A_i, B_i), \tag{10}$$

where $f(*, *)$ is mapping operator. $f(*, *) : T \times M \to \mathbb{R}$ is a two-variable single valued function defined on $T \times M$.

The seesaw mapping is as follows.

Step 1 (normalize each variable).

$$\begin{aligned} A^{\text{norm}} &= \mathscr{F}(A), \\ &\text{then } A_i^{\text{norm}} = \mathscr{F}(A),\ A_i^{\text{norm}} \in [0, 1] \\ B^{\text{norm}} &= \mathscr{F}(B), \\ &\text{then } B_i^{\text{norm}} = \mathscr{F}(B),\ B_i^{\text{norm}} \in [0, 1], \end{aligned} \tag{11}$$

where the function $\mathscr{F}(*)$ is to normalize the variables.

If two variables are measured in different units, various normalization methods are used to map the data of each variable to a range of 0~1. The dimensional expressions are transformed into dimensionless expressions, that is, scalar quantity. In many scenarios, the dimensions of the two variables are different, and it is difficult to compare the size directly, so the two variables need to be converted into scalar quantities.

Step 2 (perform coordinate transform). The developing vector of a variable is taken by a series of values in which the variable varies with T.

Definition 6 (positive/negative vectors). For the variable that plays a positive role in the combined effect, one sets it as a positive vector. The other vector is a negative vector.

Definition 7 (equilibrium factor). When the ratio of the normalized variables reaches a constant, the effect of two variables on a third variable is always the same; that is, the two variables are in equilibrium. This constant is defined as the equilibrium factor. It is denoted by γ. Obviously, if $A_i^{\text{norm}} / B_i^{\text{norm}} = \gamma$, then $G_i(A_i, B_i) = C$. C is a constant.

Definition 8 (seesaw axis). It is a coordinate axis measuring the degree of the seesaw effect in two variables. The equilibrium state is the origin.

In Figure 2, a rectangular coordinate system is made up of the developing vectors of two variables. Draw a ray with an angle of $\tan^{-1}\gamma$ in the counterclockwise direction where the coordinates of the positive vector are located. We can take one point at the rays and make a vertical line. The pedal can be used as the origin of axis. The direction of this vertical line, which is near the positive vector, is the positive direction of the axis. The other is the negative direction. This vertical line is used as the seesaw coordinate of the two variables. The number of negative directions to the origin is negative, and the number from the origin to the positive is a positive number.

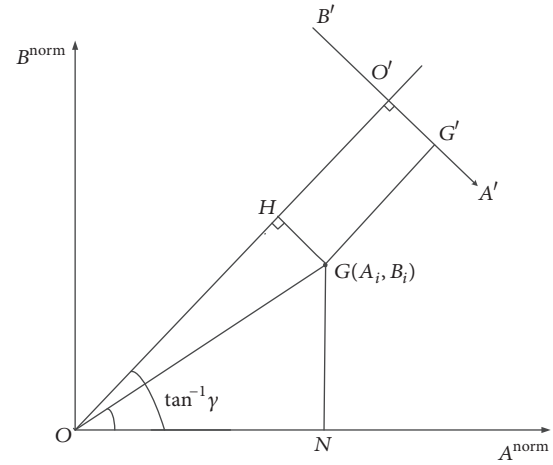


Figure 2: The diagram of the proof for saw transform.

Theorem 9. *The values of two variables A and B at a point i in T are, respectively, normalized. Let A^{norm} be the positive vector and B^{norm} be a negative vector. The value of $G_i^{norm}(A_i^{norm}, B_i^{norm})$ in the seesaw axis is*

$$\begin{aligned} g_i^{norm}(A_i^{norm}, B_i^{norm}) &= f(A_i^{norm}, B_i^{norm}) \\ &= \sqrt{A_i^{norm2} + B_i^{norm2}} \\ &\quad \times \sin\left(\tan^{-1}\gamma - \tan^{-1}\frac{B_i^{norm}}{A_i^{norm}}\right), \\ &\quad g_i^{norm}(A_i^{norm}, B_i^{norm}) \in \mathbb{R}. \end{aligned} \tag{12}$$

Proof. Assume that two variables A and B, which are normalized, can be mapped into an Euclidean space. Let OA^{norm} be horizontal ordinate and OB^{norm} be vertical ordinate. It is shown in Figure 2 that the ith state of A and B $G_i(A_i, B_i)$ is mapped as the point G in the Euclidean space. $\angle AOH = \tan^{-1}\gamma$. Take a spot O' at the ray OH. Draw a vertical line $A'B'$ from the O' pedal. The direction close to the coordinate A is chosen as the positive direction. Finally, $O'A'$ can be used as the seesaw axis. The projection of $G_i(A_i, B_i)$ on the axis $O'A'$ is G_i'. Then, the value $g_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}})$ of $G_i(A_i, B_i)$ in the seesaw axis is the length of $O'G'$.

$$\because \tan\angle NOG = \frac{NG}{ON} = \frac{B_i^{\text{norm}}}{A_i^{\text{norm}}}$$

$$\therefore \angle NOG = \tan^{-1}\frac{B_i^{\text{norm}}}{A_i^{\text{norm}}}$$

$$\because \sin\angle HOG = \sin\left(\tan^{-1}\gamma - \angle NOG\right) = \frac{HG}{OG}$$

$$\because OG = \sqrt[2]{ON^2 + NG^2} = \sqrt{A_i^{\text{norm2}} + B_i^{\text{norm2}}}$$

$$\therefore HG = OG \times \sin\angle HOG$$

$$= \sqrt{A_i^{\text{norm2}} + B_i^{\text{norm2}}} \times \sin\left(\tan^{-1}\gamma - \angle NOG\right)$$

$$= \sqrt{A_i^{\text{norm2}} + B_i^{\text{norm2}}} \times \sin\left(\tan^{-1}\gamma - \tan^{-1}\frac{B_i^{\text{norm}}}{A_i^{\text{norm}}}\right)$$

$$\therefore \text{HG} = \text{OG} \times \sin\angle\text{HOG} \quad (13)$$

$$= \sqrt{A_i^{\text{norm2}} + B_i^{\text{norm2}}} \times \sin\left(\tan^{-1}\gamma - \angle NOG\right)$$

$$\because \angle OHG = \angle HO'G' = \angle O'G'G = \frac{\pi}{2}$$

$$\therefore HG = O'G'$$

$$\therefore O'G'$$

$$= \sqrt{A_i^{\text{norm2}} + B_i^{\text{norm2}}} \times \sin\left(\tan^{-1}\gamma - \tan^{-1}\frac{B_i^{\text{norm}}}{A_i^{\text{norm}}}\right)$$

$$\therefore g_i^{\text{norm}}\left(A_i^{\text{norm}}, B_i^{\text{norm}}\right) = \sqrt{A_i^{\text{norm2}} + B_i^{\text{norm2}}} \times \sin\left(\tan^{-1}\gamma - \tan^{-1}\frac{B_i^{\text{norm}}}{A_i^{\text{norm}}}\right).$$

□

Property 10. $g_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}}) \in [-1, 1]$.

Proof. The extended seesaw axis $B'O'A'$ in Figure 2 is shown in Figure 3. The intersection of the seesaw axis and the horizontal ordinate OA^{norm} is A'. The intersection of the seesaw axis and the horizontal ordinate OB^{norm} is B'. At the point A', the value of axis reached the maximum, and then $A_i^{\text{norm}} = A^{\text{norm}}$ and $B_i^{\text{norm}} = 0$, such that $g_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}}) = 1$. At the point B', the value of axis is minimum, and then $B_i^{\text{norm}} = B^{\text{norm}}$ and $A_i^{\text{norm}} = 0$, such that $g_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}}) = -1$:

$$\therefore g_i^{\text{norm}}\left(A_i^{\text{norm}}, B_i^{\text{norm}}\right) \in [-1, 1]. \quad (14)$$

At the maximum or minimum point, the seesaw mapping is reduced to a size comparison of the values of a variable.

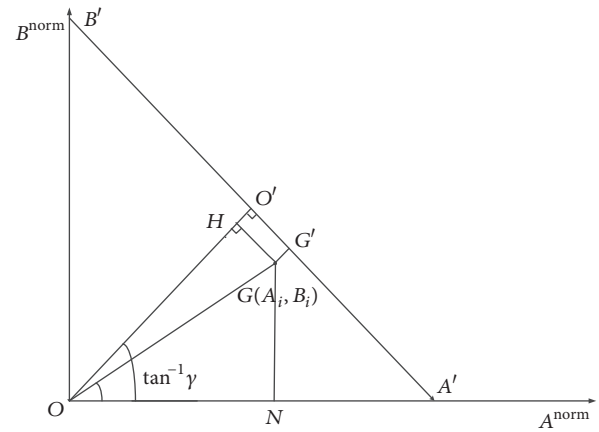


Figure 3: The diagram of the proof for saw transform property.

In this paper, it is assumed that the contribution of the common effect of D and T to the weight is not 0 in the process of weight calculation. If $\tan^{-1}\gamma = \tan^{-1}(B_i^{\text{norm}}/A_i^{\text{norm}})$, then $\sin(\tan^{-1}\gamma - \tan^{-1}(B_i^{\text{norm}}/A_i^{\text{norm}})) = 0$, and there is $g_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}}) = 0$. It requires the further evolution of Theorem 9. □

Theorem 11. *Positive values of the common contribution* $G_i^{norm}(A_i^{norm}, B_i^{norm})$ *of two variables A and B to the third variable are*

$$\widehat{g}_i^{norm} = e^{\sqrt{A_i^{norm2} + B_i^{norm2}} \times \sin(\tan^{-1}\gamma - \tan^{-1}(B_i^{norm}/A_i^{norm}))}. \quad (15)$$

Proof. ∵ The domain of the exponential function $f(x) = e^x$ is R, the range is $f(x) > 0$, and in the domain $f(x)$ increases with the increase of x.

∵ $G_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}})$ reflects the comparative relationship between the common effects of two variables at different points in T. Since the exponential function $f(x) = e^x$ is an increasing function, it does not affect the comparative relationship between them. At the same time, it can ensure that $G_i^{\text{norm}}(A_i^{\text{norm}}, B_i^{\text{norm}})$ is more than 0.

∴ The domain of the exponential function $f(x) = e^x$ can be used as a quantized function of the common effect of two variables A and B. □

4.2. Computation of the Weight Based on the Seesaw Mapping. The probability of the occurrence of an abnormal event and the effect of an abnormal event in each sensing grid are always different, so the importance of each sensing grid is different. Let the weight of the (i, j)th sensing grid be denoted by $W_{i,j}$. The matrix of the sensing grids' weights is W'.

The weight $W_{i,j}$ of each sensing grid is affected by many factors. In this paper, we list the three main factors: the probability of an abnormal event occurrence $PL_{i,j}$, the cost of damage resulting from an abnormal event, that is, the extent of the impact $D_{i,j}$, and the tolerance time of an abnormal event, that is, the period time over which the monitored object exhibits an abnormal event to cause great losses $T_{i,j}$. These three quantities of monitoring objects can be estimated by fuzzy comprehensive evaluation method or by Delphi

method. The larger the value of $D_{i,j}$ of a sensing grid is, the smaller the value of $T_{i,j}$ is and the more important the sensing grid is. $D_{i,j}$ plays a positive role in $W_{i,j}$. $T_{i,j}$ plays a negative role in $W_{i,j}$. $PL_{i,j}$, which is the premise of the combined effect of $T_{i,j}$ and $D_{i,j}$, plays a decisive role in $W_{i,j}$. Therefore, $PL_{i,j}$ is directly proportional to $W_{i,j}$ and is proportional to the combined effect of $T_{i,j}$ and $D_{i,j}$. Now, we need to study the extent of the combined effect of $T_{i,j}$ and $D_{i,j}$ on $W_{i,j}$ in each sensing grid. However, the units of measurement for $T_{i,j}$ and $D_{i,j}$ are different. The changes in $T_{i,j}$ and $D_{i,j}$ are independent of each other. These satisfy the definition of the seesaw model. We can use the seesaw mapping to compute the weights.

Step 1. Using the maximum range formula, $D_{i,j}$ is normalized into $D'_{i,j}$:

$$D'_{i,j} = \frac{D_{i,j} - D_{\min}}{D_{\max} - D_{\min}}. \tag{16}$$

Step 2. Using the maximum range formula, $T_{i,j}$ is normalized into $T'_{i,j}$:

$$T_{i,j} = \frac{T_{i,j} - T_{\min}}{T_{\max} - T_{\min}}. \tag{17}$$

Step 3. Compute the weight $W_{i,j}$ of each sensing grid.

The role of $W_{i,j}$ is to calculate the criticalness of a sensing grid relative to other grids, so we compute the weight using the normalized abnormal event tolerance time $T'_{i,j}$ and the normalized extent of the impact of the abnormal events $D'_{i,j}$. If $D'_{i,j}$ goes up in increments and $T'_{i,j}$ goes toward decreasing direction, it is beneficial to $W_{i,j}$. $T'_{i,j}$ and $D'_{i,j}$ are in a state of seesaw. The extent of the combined effect of $T'_{i,j}$ and $D'_{i,j}$ can be computed using the seesaw mapping.

The smaller the value of $T'_{i,j}$ is, the greater the value of $D'_{i,j}$ is and the more important the sensing grid is. $D'_{i,j}$ can be as the positive vector. In this work, γ is valued at 1; that is, D and T have the same criticalness. We regard that D and T have the same criticalness.

According to the seesaw mapping (Theorem 11), we know the extent of the combined effect of $T'_{i,j}$ and $D'_{i,j}$ of the (i, j)th sensing grid:

$$\begin{aligned} G_{i,j}\left(D'_{i,j}, T'_{i,j}\right) &= e^{\sqrt{D'^{\,2}_{i,j} + T'^{\,2}_{i,j}} \times \sin\left(\pi/4 - \tan^{-1}\left(T'_{i,j}/D'_{i,j}\right)\right)} \\ W_{i,j} &= G_{i,j} \times PL_{i,j}. \end{aligned} \tag{18}$$

Step 4. Normalize $W_{i,j}$ for $W'_{i,j}$:

$$W'_{i,j} = \frac{W_{i,j}}{\sum_{i=1}^{M}\sum_{j=1}^{N} W_{i,j}}. \tag{19}$$

The weights matrix W' of the monitored region is composed of many weights $W'_{i,j}$ of sensing grids.

5. The Sensing Trust Degree of WSN

As mentioned early, $R(i, j)$ is related to the total number and sensing accuracy of sensors. This paper assumes that the sensor is arranged at the center of each sensing grid. According to the sensing precision formula of single sensor node in [21], it is known that the sensing accuracy is attenuated with the increase of the distance. Assuming the position (p, q) of the grid is arranged with a sensor, the sensing accuracy formula of position (i, j) of the sensing grid is derived as

$$S_{(i,j)\to(p,q)} = \frac{1}{\left(1 + \alpha \times d_{(i,j)\to(p,q)}\right)^{\beta}}, \tag{20}$$

where $d_{(i,j)\to(p,q)}$ represents the distance between the (i, j)th and the (p, q)th sensing grid. The sensor is located in the (p, q)th sensing grid. The constants α and β are device-dependent parameters reflecting the physical features of a sensor. Generally, β ranges from 1 to 4 [17], and α is used as an adjustment parameter [16]. Sensors collect data from the sensing grid. The precision of the data in the sensing grid becomes smaller as the distance between the sensing data grid and the sensor grid increases.

When a sensing grid is equipped with a sensor, its sensing accuracy corresponding to the sensor is 1. Since $d = 0$, then $S = 1$. This denotes that the probability of sensor false positives is very small. For example, the closer you get to the camera, the clearer you see and the less likely you are to misjudge. A sensor is deployed in a sensing grid, and the sensing accuracy of the sensor grid for this sensor is 1. In this paper, we assume that sensors are deployed at the center of the sensing grid.

In the practical application of sensor networks, obstacles inevitably exist in the monitored region. These obstacles cause signal blocking and reduce sensor sensing efficiency. Here, we consider the obstruction to $S_{(i,j)\to(p,q)}$, assuming the known topography within the sensing region. If the obstacles appear in the line of sight from the grid point (i, j) to the grid point (p, q), namely, if the coordinates (x, y) of the obstacle satisfy the equations of connecting (i, j) and (p, q) line, that is, $(x - i)/(p - i) = (y - j)/(q - j)$, then $S_{(i,j)\to(p,q)} = 0$. Local blocking can also be modeled by setting the sensing probability to very small nonzero values [21].

Theorem 12. *If the coordinates (x, y) of the obstacle satisfy the equations of connecting (i, j) and (p, q) line, that is, $x \times q - x \times j - y \times p + y \times i = i \times q - j \times p$, then $S_{(i,j)\to(p,q)} = 0$.*

Proof. In actual operations, when (i, j) and (p, q) are on the same line or the same column, there is $p - i == 0$ or $q - j == 0$. It is not allowed for the denominator to be zero.

We have to transform the equation $(x - i)/(p - i) = (y - j)/(q - j)$ to $x \times q - x \times j - y \times p + y \times i = i \times q - j \times p$. □

Theorem 13. *In the model, the distance between the (i, j)th sensing grid and the (p, q)th sensor grid is*

$$d_{(i,j)\to(p,q)} = \sqrt[2]{(i - p)^2 + (j - q)^2} \times \sqrt{2} \times r_s, \tag{21}$$

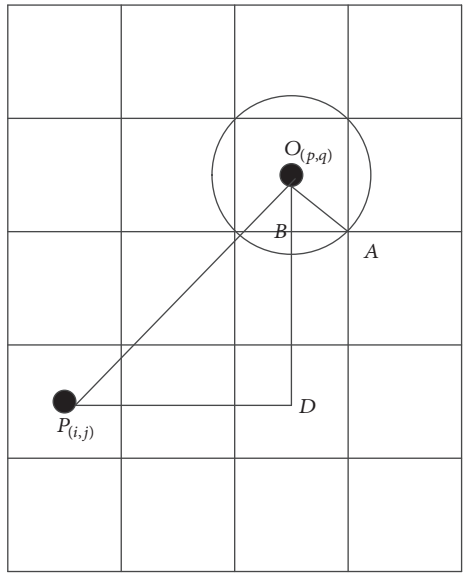


FIGURE 4: The schematic diagram of $d_{(i,j)\to(p,q)}$.

where (p, q) are the coordinates of the sensor grid in which the sensor is located.

Proof. The sensor is located at the center of the sensing grid. The sensing grid is the inscribed quadrilateral of a circle with a radius of r_s. It is shown in Figure 4 that the sensor is placed at the point O, and $d_{(i,j)\to(p,q)}$ is the distance between the grid (i, i) and the point O.

$$\begin{aligned}
\because d_{(i,j)\to(p,q)} &= OP \\
\because OA &= r_s \\
\therefore OB &= \sin\frac{\pi}{4} \times OA = \frac{\sqrt{2}}{2} \times r_s \\
\because DO &= (j-q) \times \frac{\sqrt{2}}{2} \times 2r_s \\
&= \sqrt{2} \times r_s \times (j-q) \\
\because DP &= (i-p) \times \frac{\sqrt{2}}{2} \times 2r_s \\
&= \sqrt{2} \times r_s \times (i-p) \\
\therefore d_{i,j\to p,q} &= OP = \sqrt[2]{DP^2 + DO^2} \\
&= \sqrt[2]{(i-p)^2 + (j-q)^2} \times \sqrt{2} \times r_s.
\end{aligned} \tag{22}$$

If a plurality of sensors are deployed in the sensing region, the sensing reliability degree of the (i, j)th sensing grid is

$$\begin{aligned}
R(i,j) &= \sum_{(p,q)\in P} \frac{1}{\left(1 + \alpha \times d_{(i,j)\to(p,q)}\right)^{\beta}} \\
&= \sum_{(p,q)\in P} \frac{1}{\left(1 + \alpha \times \sqrt[2]{(i-p)^2 + (j-q)^2} \times \sqrt{2} \times r_s\right)^{\beta}}.
\end{aligned} \tag{23}$$

Sensing reliability is related to whether data can be sampled by the sensor smoothly. The data in the many sensors' sensing accuracy overlapping field is read by several sensors. The reliability of the overlap field is higher because if one sensor fails, the data can also be sampled through other sensors. So, the sensing reliability is the accumulation of all sensors' sensing accuracies. □

Property 14. If $k = 0$, namely, $P = \Phi$, then $R(i, j) = 0$.

Proof. If the sensors are not deployed in the sensing matrix, then $p = \infty$, $q = \infty$.

$$\therefore R(i,j) = \sum_{(p,q)\in P} \frac{1}{k \times \left(1 + \alpha \times \infty \times \sqrt{2} \times r_s\right)^{\beta}} = 0. \tag{24}$$

□

Property 15. The range of $R(i, j)$ is $[0, 1]$.

Proof.

$$\begin{aligned}
&\because d_{i,j\to p,q} = 0 \\
&\therefore \frac{1}{\left(1 + \alpha \times d_{i,j\to p,q}\right)^{\beta}} = 1 \\
&\because d_{i,j\to p,q} = 0 \\
&\therefore \frac{1}{\left(1 + \alpha \times d_{i,j\to p,q}\right)^{\beta}} = 1 \\
&\because d_{i,j\to p,q} = \infty \\
&\therefore \frac{1}{\left(1 + \alpha \times d_{i,j\to p,q}\right)^{\beta}} = 0 \\
&\therefore 0 \le \frac{1}{\left(1 + \alpha \times d_{i,j\to p,q}\right)^{\beta}} \le 1, \\
&\because \text{The cardinality of } P \text{ is } k \\
&\therefore R(i,j) = \sum_{(p,q)\in P} \frac{1}{\left(1 + \alpha \times d_{(i,j)\to(p,q)}\right)^{\beta}} \in [0,1].
\end{aligned} \tag{25}$$

The sensing accuracy of sensors reflects the ability of sensors to collect data. The stronger the sensor, the greater the area that it covers. The coverage area of the whole WSN is the union of all sensors' coverage fields. In the overlap region,

the coverage degree takes the maximum value of all sensor accuracies. The sensing coverage degree of the (i, j)th sensing grid is

$$CO(i,j) = \max\left(\frac{1}{\left(1+\alpha\times d_{(i,j)\to(p,q)}\times 2\times r_s\right)^{\beta}}, \frac{1}{\left(1+\alpha\times d_{(i,j)\to(p,q)}\times 2\times r_s\right)^{\beta}}, \ldots, \frac{1}{\left(1+\alpha\times d_{(i,j)\to(p,q)}\times 2\times r_s\right)^{\beta}}\right)$$

$$= \max\left(\frac{1}{\left(1+\alpha\times \sqrt[2]{(i-p)^2+(j-q)^2\times\sqrt{2}\times r_s}\right)^{\beta}}, \frac{1}{\left(1+\alpha\times \sqrt[2]{(i-p)^2+(j-q)^2\times\sqrt{2}\times r_s}\right)^{\beta}}, \ldots, \frac{1}{\left(1+\alpha\times \sqrt[2]{(i-p)^2+(j-q)^2\times\sqrt{2}\times r_s}\right)^{\beta}}\right). \tag{26}$$

The total sensing coverage degree of the whole WSN is

$$CO = \sum_{i=1}^{M}\sum_{j=1}^{N} CO(i,j) = \sum_{i=1}^{M}\sum_{j=1}^{N} \max\left(\frac{1}{\left(1+\alpha\times \sqrt[2]{(i-p)^2+(j-q)^2\times\sqrt{2}\times r_s}\right)^{\beta}}, \frac{1}{\left(1+\alpha\times \sqrt[2]{(i-p)^2+(j-q)^2\times\sqrt{2}\times r_s}\right)^{\beta}}, \ldots, \frac{1}{\left(1+\alpha\times \sqrt[2]{(i-p)^2+(j-q)^2\times\sqrt{2}\times r_s}\right)^{\beta}}\right). \tag{27}$$

∵ According to Definition 2, the sensing coverage rate of the whole WSN is

$$SCO = \frac{COS}{AS}. \tag{28}$$

It is shown in Figure 4 that $AS = M\times 2\times r_s\times N\times 2\times r = 4r_s^2\times M\times N$.

$$COS = CO\times 2\times r_s\times 2\times r_s = 4r_s^2\times CO$$
$$\therefore SCO = \frac{CO}{M\times N}. \tag{29}$$

The sensing accuracy of multiple sensors overlaps in some sensing grids, and the sensing reliability of the sensing grids increases, while the effective coverage area of the whole WSN is reduced (i.e., the sensing coverage rate becomes smaller). This is because the coverage degree takes the maximum of all sensors' sensing accuracy in the overlapping areas.

Finally, the sensing trust degree of the whole WSN is defined as

$$TR\left(\Gamma\left(X\otimes W'\right)\right) = SCO\times\sum_{i=1}^{M}\sum_{j=1}^{N} W'_{i,j}\times R(i,j). \tag{30}$$

We can use the algorithms in the next section to achieve the maximum of $TR(\Gamma(X\otimes W'))$. □

6. Weight-Aware Sensor Deployment Algorithms

It is necessary to select the sensor location combination scheme whose sensing trust degree of the whole monitored region is the maximum, from the combination scheme of many sensor locations. If the same criterion is followed by each location scheme, then these combinations can be compared. This paper discusses algorithms to address the optimal sensor deployment in three cases: (1) the total number of sensors deployed in the monitored region is given; it is very hard to increase the total number of sensors; (2) the minimum sensing reliability of the sensing grid is set; (3) there are some obstacles in the sensing region.

6.1. Algorithm Notations. To describe our proposed algorithms, the following notations and message types are used.

m: the number of rows in the sensing matrix,

n: the number of columns in the sensing matrix,

k: the total number of sensors deployed in the monitored region,

α: adjustment parameter,

β: device-dependent parameter,

L: a list that stores the combinations of k numbers from 1 to $m\times n$,

g: the index of combination sequences in L,

L_g: the gth combination,

P: a list that stores a series of row coordinates and column coordinates of each sensor position corresponding to the gth combination L_g,

P^*: a list that stores a series of row coordinates and column coordinates of each sensor position corresponding to the combination whose sensing trust degree of the whole WSN is maximum,

TR: a list that stores the sensing trust degree of the whole WSN gained from each sensor deployment,

W: weight matrix,

SR: sensing accuracy degree of the whole WSN,

SCO: sensing coverage degree of the whole WSN,

$R(i, j)$: the sensing reliability degree of the (i, j)th sensing grid,

$CO(i, j)$: the sensing coverage degree of the (i, j)th sensing grid,

(i, j): the coordinate of the sensing grid,

(p, q): the coordinate of the sensing grid deployed with sensor.

6.2. Algorithm Description

Case 1. The total number of sensors deployed in the monitored region is given (see Algorithm 1).

Step 1. Generate combinatorial k numbers from 1 to $m \times n$ and store these combinations into L.

Step 2. The sensing accuracy of each sensor is added as the sensing reliability degree $R(i, j)$ of this sensing grid.

Step 3. The maximum sensing accuracy of each sensor to this sensing grid is used as the sensing coverage degree $CO(i, j)$ of this sensing grid.

Step 4. The sensing coverage degree of each sensing grid is accumulated to obtain SCO.

Step 5. The sensing reliability degree $R(i, j)$ of each sensing grid is multiplied by the weight of the sensing grid and then accumulated to obtain SR.

Step 6. Use formula (30) $TR(g) = SCO/(M \times N) \times SR$ to compute the maximum sensing trust degree of the whole WSN of the gth layout.

Step 7. Go through all the combinations in L. Put the sensing trust degree of each combination into TR. When the TR value is the maximum, the corresponding sensor position combination L_{result} is required.

Step 8. Using the function $F(\,)$, the sensor position combination L_{result} is transformed into the specific coordinates of the sensor in the sensing data matrix.

Case 2. Minimum sensing reliability $R_{\min}$ of the sensing grid is set.

In practical engineering, in order to ensure that the data of each grid is collected effectively and reliably, the minimum sensing reliability degree $R_{\min}$ of each sensing grid is often set (see Algorithm 2).

Step 1. It is necessary to determine whether $R(i, j)$ is less than $R_{\min}$ each time after $R(i, j)$ is computed. If $R(i, j)$ is less than $R_{\min}$, the current selected combination is not appropriate. There is a need to select the next combination and delete the currently inappropriate combination sequence L_g.

Step 2. If the total number of sensors is small, this will cause the sensing accuracy to be less than $R_{\min}$ in the calculation process of each sensor position combination. After going through all the combinations, the sensing accuracy is still less than $R_{\min}$, such that L is empty. There is a need to increase the number of sensors. The new combinations are generated and stored in L.

Case 3. There are some obstacles in the sensing region.

Inevitably, there will be obstacles in the monitoring area. We assume the position set of the obstacles is $OB = \{(x_m, y_n) \mid m \in \mathbb{R},\ n \in \mathbb{R},\ 0 \le m \le M,\ 0 \le n \le N\}$, where (x_m, y_n) denotes the coordinates of obstacles (see Algorithm 3).

Before $R(i, j)$ is computed, the coordinates of all the obstacles are judged whether they satisfy the line equations determined by two points at (i, j) and (p, q). If they satisfy the equation, then the value of $R(i, j)$ is zero. Otherwise, the value of $R(i, j)$ can be calculated by Algorithm 1.

7. Performance Validation

In the section, we first validate our proposed algorithm. The sensing trust degree of various sensor deployments caused by different matrix weights is compared. Then, our proposed algorithm is compared with the widely used sensor deployment algorithm MAX_AVG_COV [8].

7.1. Simulation Settings. The simulation experiment is completed on MATLAB. This paper assumes that the sensing grid is a regular square. It is assumed that each sensor has the same sensing accuracy; that is, α and β of each sensor are the same. Table 1 enumerates the common parameters of our simulation experiments.

Table 1: Simulation parameters.

Parameter description	Value
Minimum sensing radius r_s	Ten meters
Adjustment parameter α	1
Device-dependent parameter β	1

7.2. Simulation Results. In the first simulation, we study the weight's impact on the data collection effects of the WSN. The sensor deployment schemes with the same number of sensors have different sensing trust degrees of the whole WSN in the case of the different weight matrices. We choose the sensor deployment with the highest trust degree of the whole WSN as the best sensor deployment. The weight of the sensing grid in the sensing region is related to the loss cost, occurrence probability, and tolerance time of monitoring abnormal events. The weight of all the grids constitutes the weight matrix of WSN. In order to compare the different sensing trust degrees of the whole WSN caused by the

input: k, α, β
output: P^*
Generate combinatorial k numbers from 1 to $m \times n$, and store these combination into L.
for all $L_g \in L$ do
$SCO \leftarrow 0, SR \leftarrow 0$
for $i = 1$ to m do
for $j = 1$ to n do
for all $(p, q) \in P$ do

$$R(i,j) = R(i,j) + \frac{1}{\left(1 + \alpha \times \sqrt[2]{(i-p)^2 + (j-q)^2} \times \sqrt{2} \times r_s\right)^{\beta}}$$

$$CO(i,j) = \max\left(\frac{1}{\left(1 + \alpha \times \sqrt[2]{(i-p)^2 + (j-q)^2} \times \sqrt{2} \times r_s\right)^{\beta}}\right)$$

End for
$SCO = SCO + CO(i, j)$
$SR = SR + R(i, j) \times W(i, j)$
End for
End for

$$TR(g) = \frac{SCO}{M \times N} \times SR$$

End for

$$L_{\text{result}} = Lg \ when \ \max(TR(g))$$
$$P^* = F(L_{\text{result}})$$

Algorithm 1: The weight-aware sensor deployment algorithm with the given number of sensors.

Input: $R_{\min}$
Output: k
if $R(i, j) < R_{\min}$
then delete L_g
$g = g + 1$
else Perform Algorithm 1
End if
if $L == \phi \ and \ g \neq 0$
then $k = k + 1$
$g \leftarrow 0$
Perform Algorithm 1
End if

Algorithm 2: The weight-aware sensor deployment algorithm with $R_{\min}$ of each sensing grid.

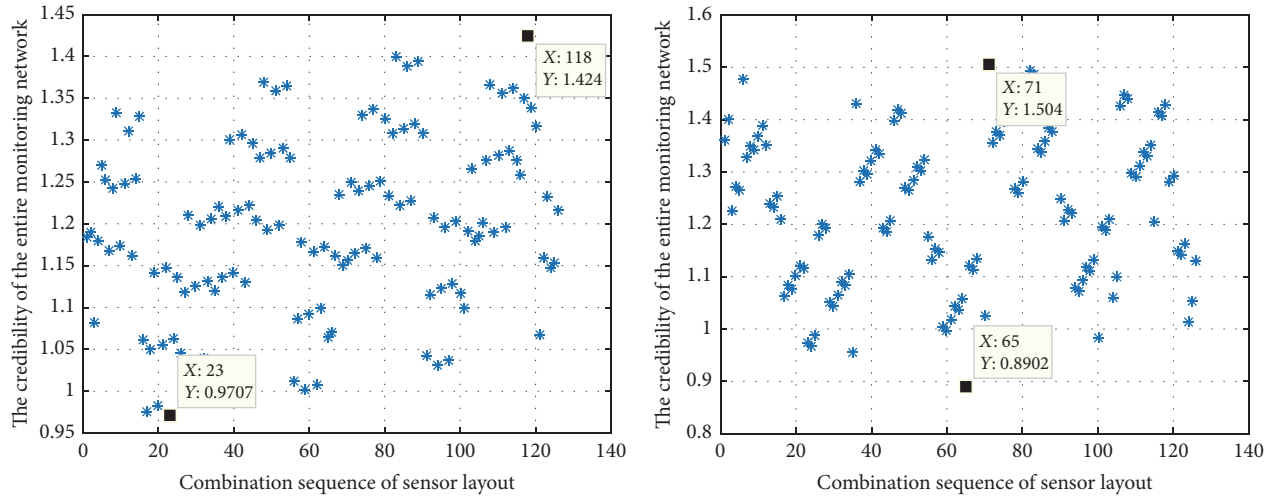
input: OB
output: $R(i, j)$
for all $(x_m, y_n) \in OB$ do
if $x \times q - x \times j - y \times p + y \times i = i \times q - j \times p$
then $R(i, j) \leftarrow 0$
End if
End for

Algorithm 3: The weight-aware sensor deployment algorithm with obstacles in some sensing grids.

different weight matrices, we randomly set the estimated parameters in Tables 2 and 3 for the weight matrices W_1 and W_2, respectively.

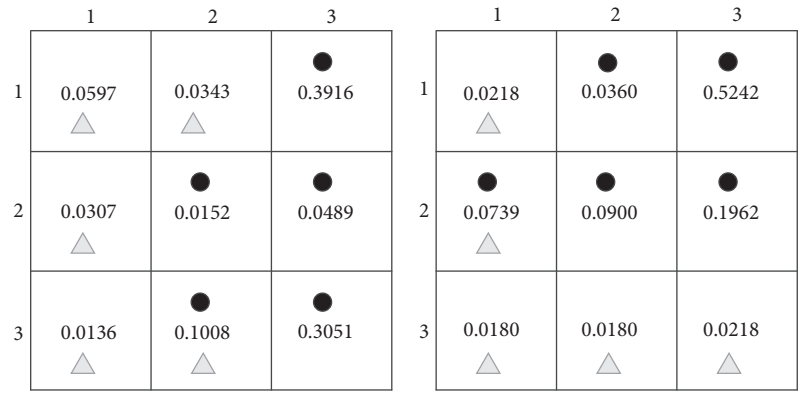
The number of sensors is set to 5. The number of abscissa grids is $M = 3$, and the number of ordinate grids is $N = 3$. We can see that the different sensing trust degrees are obtained by different weight matrices in Figure 5. When the number of sensors is 5, there are 126 sensor deployment methods, that is, 126 combinations. In Figure 5(a), the sensing trust degree of the WSN which is caused by the sensor deployment corresponding to the 118th combination is highest, up to 1.424. The sensing trust degree of the WSN which is caused by the sensor deployment corresponding to the 23rd combination is minimum, only 0.9707. In Figure 5(b), the sensing trust degree of the WSN which is caused by the sensor deployment corresponding to the 71st combination is highest, up to 1.504. The sensing trust degree of the WSN which is caused by the sensor deployment corresponding to the 71st combination is minimum, only 0.8902. The sensing trust degree of the best combination is about two times that of the worst one. Because of the weight difference, the effect of data collection on the same number of sensors is different. Finally, it is necessary to study the influence of grid weight on sensor deployment.

In order to further analyze the impact of grid weights on data collection, we compare the best deployed sensor locations with the worst locations in the case of the two different weight matrices. In Figure 6, these ● denote the sensor locations with the highest sensing degree of the whole WSN, and these △ denote the sensor locations with minimal sensing degree. In Figure 6(a), the sensor deployment coordinates with the highest sensing degree of the whole



(a) Weight matrix W_1 (b) Weight matrix W_2

FIGURE 5: The performance of different weight matrices. (a) Weight matrix W_1 and (b) weight matrix W_2.



(a) Weight matrix W_1 (b) Weight matrix W_2

FIGURE 6: Comparison of the best and worst sensor deployment.

WSN are $\{(1,3),(2,2),(2,3),(3,2),(3,3)\}$, and the sensor deployment coordinates with the minimal sensing degree are $\{(1,1),(1,2),(2,1),(3,1),(3,2)\}$. In Figure 6(b), the sensor deployment coordinates with the highest sensing degree are $\{(1,2),(1,3),(2,1),(2,2),(2,3)\}$, and the sensor deployment coordinates with the minimal sensing degree are $\{(1,1),(2,1),(3,1),(3,2),(3,3)\}$.

As is shown in Figure 6, we can see that the sensors are deployed on the edge of the sensing area and in the grids with lower weights by the sensor deployment with the minimum trust degree of the whole WSN. However, the sensors are almost deployed on the center of the sensing area and in the grids with larger weights by the sensor deployment with the maximum trust degree.

In the second simulation, we compare our proposed algorithm with MAX_AVG_COV [8]. We assume that the number of rows of the sensing matrix is 4 and the number of columns is 4. We measure the minimum number of sensors required, when miss probability threshold $M_{\min}$ (corresponding to the minimum reliability of each grid $R_{\min}$) varies from 0.1 to 0.7 with an increment of 0.1.

As we can see from Figure 7, the minimum number of sensors required decreases as the minimum reliability increases. When $M_{\min}$ is 0.1, the minimum number of sensors required by the proposed algorithm is half the number in the MAX_AVG_COV algorithm. When $M_{\min} = 0.7$, the required number of sensors in our algorithm is less than 2 in the MAX_AVG_COV algorithm. Obviously, the algorithm proposed in this paper is superior to MAX_AVG_COV algorithm.

In the third simulation, we study the presence of obstacles in the sensing area. Suppose there are 4 obstacles. The layout model of the obstacle of our proposed algorithm is shown in Figure 8. The coordinates of the obstacles are set to $OB = \{(2,2.5),(2.5,4),(2.5,1),(3.5,2)\}$.

Table 2: The estimated parameters of W_1.

The cost of damage resulting from an abnormal event D(/hundred dollars)	{11, 2, 9; 0.5, 1, 9; 0.5, 3, 8}
The tolerance time of an abnormal event T(/hours)	{2, 1, 0.5; 0.3, 3, 1; 0.1, 4, 1}
The probability of an abnormal event occurrence PL	{0.1, 0.3, 0.8; 0.9, 0.25, 0.1; 0.4, 0.6, 0.7}

Table 3: The estimated parameters of W_2.

The cost of damage resulting from an abnormal event D(/hundred dollars)	{1, 0.5, 8; 3, 10, 1; 0.5, 0.5, 1}
The tolerance time of an abnormal event T(/hours)	{1, 3, 8; 0.1, 3, 1; 0.9, 2, 1}
The probability of an abnormal event occurrence PL	{0.1, 0.2, 0.7; 0.2, 0.1, 0.9; 0.1, 0.1, 0.1}

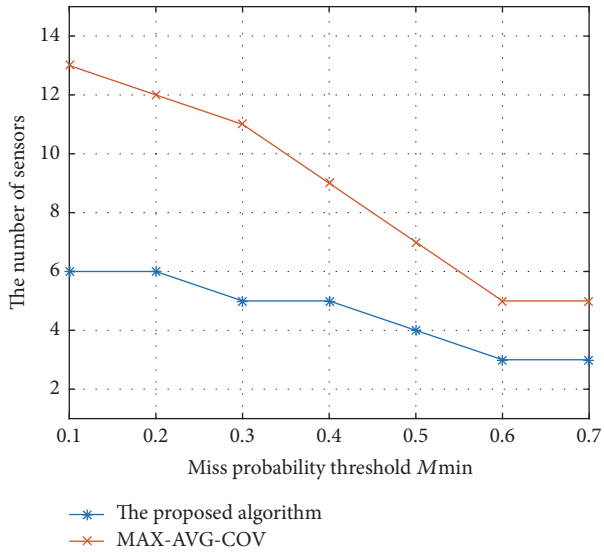


Figure 7: The comparison for minimum reliability.

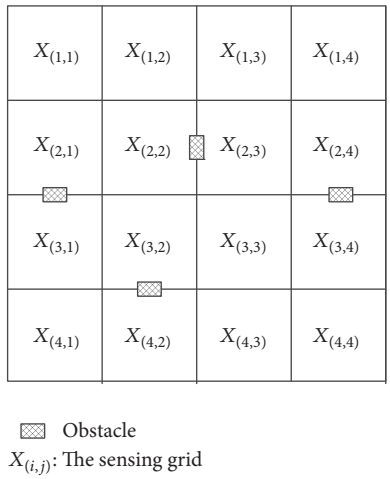


Figure 8: The layout model of the obstacle of our proposed algorithm.

The layout model of the obstacle of the MAX_AVG_COV algorithm is shown in Figure 9.

Then, $d_{10,14} = 0$, $d_{14,10} = 0$, $d_{6,7} = 0$, $d_{7,6} = 0$, $d_{8,12} = 0$, $d_{12,8} = 0$, $d_{5,9} = 0$, and $d_{9,5} = 0$, where $d_{i,j}$ denotes distance from the ith sensing grid to the jth sensing grid.

We compare the minimum number of sensors required by the two algorithms, as shown in Figure 10, when there are obstacles in the sensing region.

As we can see from Figure 10, the minimum number of sensors required varies when there are obstacles in the sensing area. The variations of the algorithms proposed in this paper are smaller. At the miss probability threshold 0.1, the minimum number of sensors required by the MAX_AVG_COV algorithm increases to 15, while that of our proposed algorithm is still 6, with a difference of 9. With the increase of miss probability threshold, the restriction of sensor deployment is getting smaller and smaller, and the minimum number of sensors required is also reduced. At the miss probability threshold 0.7, the minimum number of sensors required by our algorithm is 3, while the minimum number of sensors required by the MAX_AVG_COV algorithm is 5, with a difference of two. Since MAX_AVG_COV algorithm is a local optimum selection method, the localization of obstacles has a great influence on the sensor deployment. The proposed algorithm is a global optimal selection method; as long as the computation time is sufficient, we can always choose an optimal layout scheme, even a way of sensor deployment that

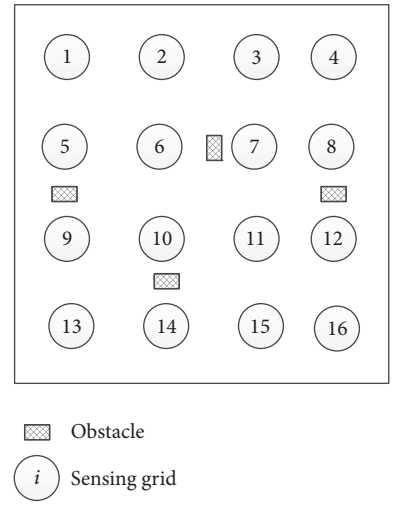


Figure 9: The layout model of the obstacle of MAX_AVG_COV algorithm.

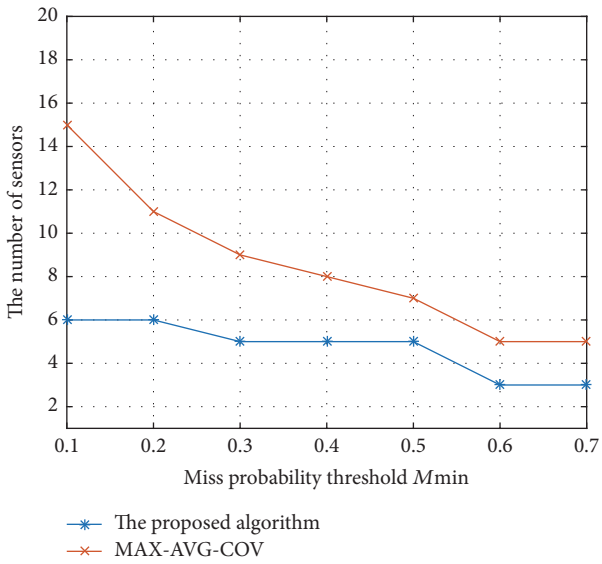


Figure 10: The comparison for the existence of obstacles.

can be chosen to avoid the effects of obstacles. From this, we can see that the algorithm proposed in this paper is obviously superior to MAX_AVG_COV algorithm.

8. Conclusion

In practical applications, especially for smart city wireless sensor network deployment, the inconsistency of data weights owing to relative criticalness in different sensing grids needs to be taken into account. In order to ensure the data collection of wireless sensor networks effectively, we propose a computing model of the sensing trust degree of the whole WSN. In this paper, the sensing region is divided into meshed grids. The weights of the sensing grids are estimated by seesaw mapping to reflect the combined effect of the variables related to abnormal events, including the probability of occurrence PC (i, j), loss cost, that is, the degree of impact $D(i, j)$, and the tolerance time $T(i, j)$. The computing model based on sensing trust degree takes into account the sensor location, the impact of the weight of sensing grid, and the obstacles in the sensing region. We study the impact of different sensor deployments on the sensing trust degree of the whole WSN. Our model can obtain as much information as possible when the number of sensors is given. Based on our proposed computing model and formulated optimization problem, we propose the sensor deployment algorithms. Our algorithms are designed for three cases: (1) the number of sensors is limited, (2) the minimum sensing precision is defined, and (3) there are some obstacles in the sensing area. In the simulation experiments, we find out that the sensing affections of different sensor deployment schemes with different weight matrices are different under the condition of the same number of sensors. When the sensing accuracy is limited, the minimum number of sensors needed by our sensor deployment algorithm is less than that of the conventional MAX_AVG_COV algorithm. In this work, the sensing grid is square. In our following research, we will study the sensor deployment scheme for irregular grids and reduce the computational complexity.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the Project of Natural Science Foundation of Hainan Province in China (Grants nos. 20166232 and 617033), the Innovative Research Projects for Graduate Students of Hainan Higher Education Institutions (Grant no. Hyb2017-06), the National Natural Science Foundation of China (Grants nos. 61561017 and 61462022), Hainan Province Major Science & Technology Project (Grant no. ZDKJ2016015), and Open Project of State Key Laboratory of Marine Resource Utilization in South China Sea (Grant no. 2016013B).

References

[1] C. Zidi, F. Bouabdallah, and R. Boutaba, "Routing design avoiding energy holes in underwater acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 16, no. 14, pp. 2035–2051, 2016.

[2] I. Vlasenko, I. Nikolaidis, and E. Stroulia, "The smart-condo: optimizing sensor placement for indoor localization," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 3, pp. 436–453, 2015.

[3] Y. Zou and K. Chakrabarty, "Uncertainty-aware sensor deployment algorithms for surveillance applications," in *Proceedings of the IEEE Global Telecommunications Conference GLOBECOM'03*, pp. 2972–2976, USA, December 2003.

[4] A. Rakavi, M. S. K. Manikandan, and K. Hariharan, "Grid based mobile sensor node deployment for improving area coverage in Wireless Sensor Networks," in *Proceedings of the 3rd International Conference on Signal Processing, Communication and Networking, ICSCN 2015*, India, March 2015.

[5] Y. Xiao and Y. Zhang, "Divide- and conquer-based surveillance framework using robots, sensor nodes, and RFID tags," *Wireless Communications and Mobile Computing*, vol. 11, no. 7, pp. 964–979, 2011.

[6] W. An, S. Ci, H. Luo et al., "Effective sensor deployment based on field information coverage in precision agriculture," *Wireless Communications and Mobile Computing*, vol. 15, no. 12, pp. 1606–1620, 2015.

[7] R. R. Mishra and L. Moharana, "Analysis of different grid types used for sensor deployment in wireless sensor network," in *Proceedings of the International Conference on Communication, Control and Intelligent Systems, CCIS 2015*, pp. 91–95, India, November 2015.

[8] S. S. Dhillon and K. Chakrabarty, "Sensor placement for effective coverage and surveillance in distributed sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference: The Dawn of Pervasive Communication (WCNC '03)*, vol. 3, pp. 1609–1614, March 2003.

[9] T. Brown, Z. Wang, T. Shan, F. Wang, and J. Xue, "Poster abstract: obstacle and connectivity aware wireless video sensor

deployment for 3D indoor monitoring," in *Proceedings of the 2nd IEEE/ACM International Conference on Internet-of-Things Design and Implementation, IoTDI 2017*, pp. 305-306, USA, April 2017.

[10] C.-Y. Chang, C.-T. Chang, Y.-C. Chen, and H.-R. Chang, "Obstacle-resistant deployment algorithms for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2925–2941, 2009.

[11] J. Long, M. Dong, K. Ota, A. Liu, and S. Hai, "Reliability guaranteed efficient data gathering in wireless sensor networks," *IEEE Access*, vol. 3, pp. 430–444, 2015.

[12] D. S. Deif and Y. Gadallah, "An ant colony optimization approach for the deployment of reliable wireless sensor networks," *IEEE Access*, vol. 5, pp. 10744–10756, 2017.

[13] F. Senel, "Coverage-aware connectivity-constrained unattended sensor deployment in underwater acoustic sensor networks," *Wireless Communications Mobile Computing*, vol. 16, no. 14, pp. 2064-2052, 2016.

[14] X. F. Xing, G. J. Wang, and J. Li, "Polytype target coverage scheme for heterogeneous wireless sensor networks using linear programming," *Wireless Communications and Mobile Computing*, vol. 14, no. 14, pp. 1397–1408, 2014.

[15] L. I. Hai-Hua, J. Fan, and L. Chen, "Application of grid method in deployment of wireless sensor networks," *Transducer & Microsystem Technologies*, 2012.

[16] L. Jun and T. Suda, "Coverage-aware self-scheduling in sensor networks," in *Proceedings of the IEEE 18th Annual Workshop on Computer Communications, (CCW '03)*, pp. 117–123, IEEE, Dana Point, Calif, USA, October 2003.

[17] S. Meguerdichian, F. Koushanfar, G. Qu, and M. Potkonjak, "Exposure in wireless ad-hoc sensor networks," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking*, pp. 139–150, ACM, July 2001.

[18] M. Fujimoto, S. Watanabe, and T. Nakatani, "Frame-wise model re-estimation method based on Gaussian pruning with weight normalization for noise robust voice activity detection," *Speech Communication*, vol. 54, no. 2, pp. 229–244, 2012.

[19] M. Xie, Y. Deng, Y. Bai, M. Huang, and Z. Hu, "Research on the pre-distribution model based on seesaw model," *Communications in Computer and Information Science*, vol. 729, pp. 201–213, 2017.

[20] W.-L. Guo and Z.-Z. Xing, "Constraints on the tritium beta decay and the neutrinoless double beta decay in the minimal seesaw model," *High Energy Physics and Nuclear Physics*, vol. 30, no. 8, pp. 709–713, 2006.

[21] L. Li, Q. Li, Z. Yujia, and Z. Baohua, "A highly efficient algorithm for node placement," *Journal of University of Science & Technology*, vol. 35, no. 3, pp. 411–416, 2005.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4796535, 12 pages
https://doi.org/10.1155/2018/4796535

Research Article

Augmenting High-Performance Mobile Cloud Computations for Big Data in AMBER

Muhammad Munwar Iqbal,[1] Muhammad Ali,[1] Mai Alfawair,[2] Ahsan Lateef,[3] Abid Ali Minhas,[4] Abdulaziz Al Mazyad,[4,5] and Kashif Naseer[6]

[1]*Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan*
[2]*Prince Abdullah Bin Ghazi Faculty of Information Technology, Al-Balqa' Applied University, Salt, Jordan*
[3]*Department of Computer Science, University of Agriculture, Faisalabad, Pakistan*
[4]*College of Computer and Information Systems, Al Yamamah University, Riyadh, Saudi Arabia*
[5]*King Saud University, Riyadh, Saudi Arabia*
[6]*Department of Computer Engineering, Bahria University, Islamabad, Pakistan*

Correspondence should be addressed to Abid Ali Minhas; a_minhas@yu.edu.sa

Received 30 August 2017; Accepted 30 January 2018; Published 2 April 2018

Academic Editor: Syed H. Ahmed

Big data is an inspirational area of research that involves best practices used in the industry and academia. Challenging and complex systems are the core requirements for the data collation and analysis of big data. Data analysis approaches and algorithms development are the necessary and essential components of the big data analytics. Big data and high-performance computing emergent nature help to solve complex and challenging problems. High-Performance Mobile Cloud Computing (HPMCC) technology contributes to the execution of the intensive computational application at any location independently on laptops using virtual machines. HPMCC technique enables executing computationally extreme scientific tasks on a cloud comprising laptops. Assisted Model Building with Energy Refinement (AMBER) with the force fields calculations for molecular dynamics is a computationally hungry task that requires high and computational hardware resources for execution. The core objective of the study is to deliver and provide researchers with a mobile cloud of laptops capable of doing the heavy processing. An innovative execution of AMBER with force field empirical formula using Message Passing Interface (MPI) infrastructure on HPMCC is proposed. It is homogeneous mobile cloud platform comprising a laptop and virtual machines as processors nodes along with dynamic parallelism. Some processes can be executed to distribute and run the task among the various computational nodes. This task-based and data-based parallelism is achieved in proposed solution by using a Message Passing Interface. Trace-based results and graphs will present the significance of the proposed method.

1. Introduction

Big data challenges are opportunities for the academic as well as for the industry. Data processing and analytics of high-performance computing machinery provide the possibilities for designing complex business models, highly scalable and reconfigurable systems. The main interest of this research is to provide researchers with desktop-based machines or virtual machines, having the capacity of massive processing to run the jobs that require high-end hardware like supercomputers. Cloud computing can be used to fulfil the requirements for the computation-intensive jobs.

Cloud computing running programs that run across homogeneous platforms on multiple nodes consists of multiple virtual machines. MPI provides parallel computing software infrastructure, task-based and data-based parallelism. This solution is cost-effective and efficient as compared to previous ones. Cloud computing data security and privacy are enhanced by using the Single Sign-On approach. This approach uses identity management techniques like OAuth,

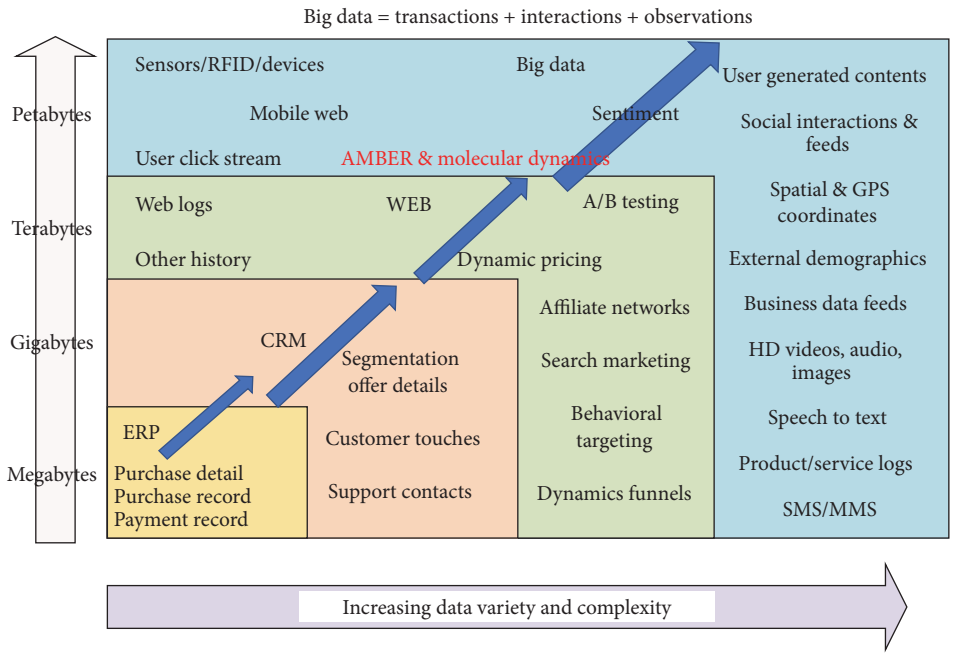


Figure 1: Multimedia big data particular levels [4, 5].

Table 1: Big data challenges with cloud computing.

Big data/cloud computing	Elasticity	Pooled	On-demand	Self-service	Pay-as-you-go
Volume		x			x
Velocity	x		x		
Variety	x	x		x	
Veracity				x	x
Value	x		x		x

OpenID, and SAML for the enhancement of cloud users' security and privacy [1].

The main challenges are content management, security, transmission, information retrieval, and mining in multimedia databases [2]. While redefining the possibilities of molecular dynamics and AMBER [3], cloud computing has engaged big data and enlightened potential solutions for various digital earth problems in geoscience and relevant domains such as social science, astronomy, business, and industry as shown in Figure 1.

In this research, energy calculation of AMBER force fields of molecular dynamics has been selected to prove the concept. This method is used to have an estimation of molecular structures and energies for confirmations of physical properties. The Internet is the primary source of the multimedia data produced on a large scale and freely available.

The features of cloud computing and their utilization to support characteristics of big data are summarized Table 1.

Molecular dynamics is a computer-based recreation of the developments of particles considering N-number of atoms as an N-body reenactment. The atoms and particles associated with each other for a characterized time frame, accordingly, furnish us with a perspective of the translational developments of molecules. In the regular adaptation of reproductions, Newton's conditions of movement for an arrangement of N-communicating particles are unravelled numerically to watch the directions of iotas and atoms. The strengths between the particles and potential vitality are ascertained by utilizing any one sort of compelling fields of subatomic mechanics. It is a procedure of hypothetical physical science that was considered in the decade of 1950s, yet it is received right now for synthetic and physical investigation, materials science examination, and the displaying of biomolecules [6].

2. 7 Vs of Big Data

The cloud and IoT data are growing exponentially, that is, previously in kilobytes, megabytes, and gigabytes. The big data that was previously comprised of 4 Vs, but now it is pretty well changed into 7 Vs data, which are velocity, variety, variability, veracity, visualization, value, and volume. Molecular dynamics contains the massive volume of unstructured and structured data referred to as big data. The term "big data [3]" is originated from the web companies and they had to query the loosely structured large shared data. Big data is composed of seven principal terms, which signify

its importance. These terms are velocity, variety, variability, veracity, value, volume, and visualization. Velocity shows the speed of data generation, processing, and storage. It depicts the data that can be accessed with enough speed.

Variety is the most significant challenge in big data and is a range of structured, semistructured, and unstructured data. These data sources have many data types like XML, audio, video, SMS, and sensor data. The organization of the data is the not an easy task; mainly when the data types are different, data format is heterogenetic; multifactor data values and probabilistic data are also available for processing. Variability is different from the variety and veracity like the taste of six different blends of coffee at the shop. Since data meaning is continually changing, metadata also changes accordingly on homogenization of data.

Veracity means to ensure the data that is required for processing is accurate and keep away bad, dirty data from collecting into systems for execution, like entering false and incorrect names and wrong contact information into databases. Visualization is an essential and critical element of big data. The sophisticated data visualization uses graphs, charts, modeling tools, and techniques with reported chockfull formulae and numbered values.

Value is the most effective element of the big data after putting whole effort and addressing the volume, velocity, variety, variability, veracity, and visualization. The organization must be getting values from data which is beneficial for setup. Big data now possibly define the data in zettabytes and yottabytes. The volume of data is currently increasing in gigabytes per minutes.

The big data are retrieved from heterogeneous and multiple resources, as well as their ingredients. Therefore the complexity of data is the core concern. The big data must be cleansed, matched, transformed, and linked into chosen formats before processing [7]. The other examples are Facebook and Twitter, which are generating data through user social interaction [8].

There are numerous techniques to figure out the energies and constraints; one path is by computing the traditional compelling fields or a quantum hypothesis at close level. The number count and the speeds of particles are introduced with starting self-assertive qualities that match the aggregate entirety of dynamic energies of the framework, which, thus, is further managed by the required temperature of reproduction. By having the estimations of the drive on every molecule, we can ascertain the increasing speed of every particle in the entire framework. Count of the integral of Newton's conditions of movement then decides a direction that depicts the quickening, position, and speed of the iotas and atoms which shift with time. The standard estimations of properties can be computed by keeping in view these directions. It is known as a deterministic strategy for figuring. In this technique, the speeds and positions of every single iota are utilized. By running the reenactment, these qualities can be ascertained for any moment of time. The time moment can browse the future or the past. Subatomic elements recreations are tedious and computationally costly if accomplished for expensive frameworks [9].

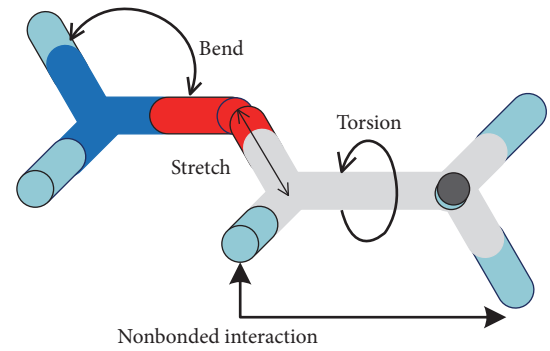


Figure 2: Force fields bonded and nonbonded interactions.

Molecular dynamics reproduction is one of the fundamental techniques in the domain of subatomic rebuilding and is utilized for subatomic examination [10, 11]. Atomic reenactments give us dependable association in force fields. They help us to learn the plainly visible thermodynamic properties of particles, for example, compressibility, inward vitality, weight, warm extension, particular warms, pliable and shear modulus that can be assessed. It is finished by utilizing the infinitesimal level data acquired by the aftereffects of reenactments [10].

Specific commands are utilized to extricate force field information from the prototypical model. It is also used for the vitality expression or capacity calculation of the particular model. This specific vitality expression is the condition that communicates the possible vitality on the surface of a specific model. This is portrayed as an element of its nuclear directions. The sum of valence or reinforced and nonfortified associations can be communicated as the potential vitality of a framework. The corner ordinarily represents the energies required in valence connections to corner terms. These energies are named as bond extend term (E_{bond}), valence point twist term (E_{angle}), dihedral edge torsion term (E_{torsion}), and reversal ($E_{\text{inversion}}$) terms. These vitality terms are a piece of all constraint fields vitality estimations for covalent frameworks as shown in Figure 2.

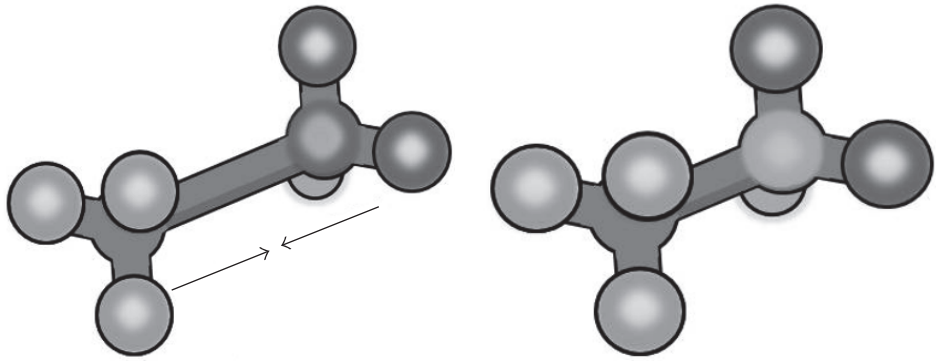
The E_{val} is represented the by

$$E_{\text{val}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{inv}}. \quad (1)$$

The interactions' energy of the nonbonded atoms is calculated by considering van der Waals (E_{vdW}), electrostatic (E_{Coulomb}), and hydrogen bond (E_{hbond}) terms as shown in

$$E_{\text{nonbond}} = E_{\text{Coulomb}} + E_{\text{vdW}} + E_{\text{hbond}}. \quad (2)$$

AMBER is a method with force fields energy calculation for molecular dynamics. The equation for this type of force fields energy calculation defines the total potential energy of the molecules. The AMBER in minimization of the bond stretching energy is shown in Figure 3. This force is dependent concerning the position and its potential derivative of the equation to calculate this energy is called a functional form of AMBER.



(a) (b)

Figure 3: AMBER in minimization of the bond stretching energy.

The AMBER force fields functional form is described as the equation mentioned below [12, 13]:

$$V\left(r^N\right) = \sum_{\text{bonds}} k_b \left(l - l_0\right)^2 + \sum_{\text{angles}} k_a \left(\theta - \theta_0\right)^2 + \sum_{\text{torsions}} \frac{1}{2} V_n \left[1 + \cos\left(n\omega - \gamma\right)\right] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ \epsilon_{i,j} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}, \tag{3}$$

where k_a is the Angle Constant; θ is the angle bend; ϵ is the well depth; k_b is the bond constant; ω is the torsion angle; V_n is the torsion constant; l_0 is the natural bond length; l is the bond length; q_i is the charge on i; q_j is the charge on j; and γ is the phase factor. The expression on the right-hand side of AMBER force fields functional form is described in detail as follows: [14, 15]

(i) Summation over bonds is the first term, used to calculate the energy of the atoms that are bonded covalently. This harmonic force is also known as ideal spring force. It gives us good approximation close to the bond length equilibrium, but this force becomes poorer as the distance between atoms increases.

(ii) Summation over angles is the second term, used to present the energy, which is due to the geometrical placement of electron in orbitals involved in covalent type bonding.

(iii) Summation over torsions is the third term, used for the representation of the energy due to twisting a bond as a result of the bond order, double bonds, and neighbouring electrons in lone pairs or bonds. A solo bond can have one or more than one term; for example, the total torsional energy can be expressed as a Fourier series.

(iv) Double summing for i and j is the fourth term, used to express the nonbonded energy among all the pairs of atoms. This term can further be decomposed into two terms: the first term of the summation is van der Waals forces and the second term of the summation is electrostatic energies.

The equilibrium distance (r_{0ij}) with depth (ϵ) is used for calculating the van der Waals energy, and equilibrium distance is r_{0ij} ensured by the factor of 2, in the calculation. Terms of σ, where $r_{0ij} = 2^{1/6}(\sigma)$, are sometimes used for the reformulation of energy. It may be used for example, in the implementation of the soft-core potentials [16].

The researcher reported in this investigation that the computation-intensive architecture was used free from geolocations for the computation of the Assisted Model Building with Energy Refinement (AMBER) with the force fields calculations for molecular dynamics. The proposed scheme is also useful for providing mobility to the calculation of the computation-intensive jobs, where Internet connectivity is not possible. The proposed methodology is providing software based instead of previously proposed solutions, which are costly and hardware based and setup is more critical. This proposed solution is composed of open source platform which minimized the computation cost.

3. Literature Review

OpenCL is embraced by Apple, Intel Xeon Phi, Qualcomm, Samsung, Vivante, Advanced Micro Devices (AMD), Altera, Nvidia, and ARM Holdings [17, 18]. The force fields figuring is an experimental computation technique proposed to give evaluations of structures and energies for adaptations of particles. The enhancement and conformities required to run the computation-intensive jobs need a testbed composed of very costly hardware. The investigation is also helpful for the exascale calculation in future [5]. The technique depends on the presumption of "common" bond lengths and points, deviation from which prompts strain, and the presence of torsional connections and appealing or potentially terrible van der Waals and dipolar powers between nonfortified molecules [19]. The strategy is called experimental drive field gullible.

Molecular dynamic (MD) progression is a computer recreation of physical developments of particles and atoms with regard to N-body reproduction [20]. The iotas and particles are permitted to collaborate for a timeframe, giving a perspective of the movement of the molecules. In the most widely recognized adaptation, the directions of atoms and iotas are numerically dictated by measuring Newton's conditions of movement for an arrangement of interfacing particles, where subatomic mechanics characterizes drives between the atomic particles and potential vitality compelling fields. The strategy was initially considered inside theoretical physical science in the late 1950s, yet is connected today for the most part to synthetic material science, materials science, and the displaying of biomolecules [21].

Electrostatic energy can be used here to assumes single point charge in the equation. It means that a single point charge can represent both electrons and protons in an atom or in the case of parameter sets that employ point charges and electron lone pairs. If we consider a conventional molecular dynamic (MD) simulation, the most processor-hungry job is the calculation of the force field energies as a function of the internal coordinates of atoms and molecules that are under considerations. In the evaluation of that energy, the noncovalent and nonbonded energy calculations are a most computationally intensive part. In the view of complexity, big "O" notation of the MD calculations is scaled by $O(n^2)$. This expression is valid only if we explicitly consider all pair-wise electrostatic and van der Waals interactions for force fields energy calculation [22].

One may assess the energies and powers utilizing traditional drive fields or a chose level of quantum hypothesis. By and by, the iotas are allocated beginning speeds that comply with the aggregate dynamic vitality of the framework, which, like this, is directed by the carved reproduction temperature. From the information of the compel on every molecule the increasing speed of every particle in the framework can be decided. Incorporation of the conditions of movement then yields a direction that depicts the positions, speeds, and increasing speeds of the particles as they differ with time. From this direction, the normal estimations of properties can be resolved. The technique is deterministic; once the positions and speeds of every molecule are known, the condition of the framework can be anticipated, whether later on or in the past. Subatomic elements reproductions can be tedious and computationally costly [23]. Atomic progression (MD) recreation is one of the essential strategies in the subatomic reproduction world [24]. Through atomic reenactments with reliable communication compelling fields, the naturally visible thermodynamic properties, for example, weight, inner vitality, warm development, compressibility, tractable and shear modulus, particular warms, can be assessed by utilizing the minute level data produced through recreations [25].

Mixture parallel computing is one approach for addressing the issues of processor-hungry molecular dynamics recreation. Its columns depend on basic heterogeneous processing framework, climate CPU/GPU blend, or Intel's Xeon Phi coprocessor. Heterogeneous figuring refers to frameworks that use more than one sort of processors. These are multicenter frameworks that pick up execution by including centers, as well as by fusing specific preparing abilities to handle specific assignments. Heterogeneous System Architecture (HSA) frameworks use various processor sorts (typically CPUs and Intel Phi coprocessors), for the most part on similar silicon pass on, to give you the best of both spaces. First is vector handling; aside from its outstanding parallel preparing abilities, it can likewise perform numerically concentrated calculations on huge information sets. Second is CPUs that can run the practical framework and perform conventional serial undertakings [26, 27].

Seven frameworks on the top 500 rundowns were at that point utilizing the Intel chip. The top supercomputers, Intel's quickening agent share is still a minute, machines utilizing Nvidia's GPUs dwarf, those with Intel's coprocessors 31 to 11. Tianhe-2 has significantly affected the scene. In the event of counting up all the peta-tumbles on the rundown, "the collected execution conveyed by Intel Xeon Phi coprocessors is currently greater than the execution conveyed by GPU quickening agents," says Intel representative Radoslaw Walczyk. "It is a major win," says Sergis Mushell, an examiner at the innovation inquiries about firm Gartner. Intel's chips contain up to 61 centers and are assembled utilizing the organization's 22-nanometer fabricating process, which is an era in front of the opposition. The organization says its coprocessors have a couple of focal points over GPGPUs: they can work freely of CPUs and they do not require a unique code to program [28].

Intel gives C and C++ languages compiler to Intel Xeon Phi coprocessor. OpenCL is summed up design for programming accessible on all top-of-the-line gadgets (mobile GPUs, AMD GPUs, Intel gadgets). C and C++ languages are recommended to meet the necessities of hybrid parallel computing as its streamlining instruments, for example, V Tune is likewise accessible by Intel to upgrade our application [27, 29, 30].

An amazing drive field, situated on the atomistic depiction of the silicon dioxide statement on a melted silica substrate, has been produced and connected to the subatomic element recreation with the GROMACS bundle. The legitimacy of the created recreation approach is checked utilizing nuclear bunches comprising up to 106 molecules and having trademark measurements of up to 30 nm. The C and C++ languages advancement apparatus is bundled with Intel Composer XE. This bundle incorporates the libraries, debuggers, and the compilers. It has instruments to fabricate the offload and cross-incorporated form of the hotspot for Intel Xeon Phi coprocessors. It perceives compiler-upheld punctuation and its utilization concerning Intel Xeon Phi [30, 31]. The OpenCL Applications XE 2013 R2 uses Intel SDK for a product advancement environment for OpenCL applications for Intel Xeon processors and Intel Xeon Phi™ coprocessors [31]. The SDK gives documentation, improvement devices, and the OpenCL 1.2 runtime for Intel Xeon Phi coprocessors and Intel Xeon processors. The Intel SDK for OpenCL Applications XE 2013 for Linux is utilized. The SDK bolsters both the Intel Xeon server and Intel Xeon Phi coprocessor.

The Intel Xeon Phi coprocessor is the main item given, Intel Much Integrated Core Architecture (Intel MIC engineering), and it targets HPC sections, for example, oil

Table 2: Parameters required to calculate bond stretch terms and angle bend terms.

Sr. No.	Parameter	Description
(1)	"forceConstant"	Constant variable used in calculation
(2)	"equilibrium distance."	Distance between two atoms
(3)	"angleBendTerms [ix5]"	All the four values of angle bend terms according to functional form
(4)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array
(5)	"dihedralAngleTerms [ix7]"	All the four values of dihedral angle terms according to functional form
(6)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array
(7)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array
(8)	"nonBondedTerms [ix4]"	All the four values of nonbonded terms according to functional form
(9)	"coulomb factor."	It is calculated using permittivity and permittivity scale and a constant value 332
(10)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array
(11)	"bond length."	The natural bond length
(12)	"energy"	The partial value of energy due to bond stretch terms is calculated and stored in it
(13)	"angleBendTerms [ix5]"	All the four values of angle bend terms according to functional form
(14)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array
(15)	"theta deg"	It is calculated in this function from coordinates
(16)	"term"	Intermediate value to calculate energy due to angle bend
(17)	"energy"	The partial value of energy due to angle bend terms is calculated and stored in it

investigation, logical research, money detailed examinations, and atmosphere reproduction, among numerous others. Intel MIC design joins various Intel CPU centers onto a single chip. Engineers keen on programming these centers can utilize standard programming techniques. Similar OpenCL source code composed of Intel Xeon processor can be reused with minor changes on the Intel Xeon Phi coprocessor [32, 33].

4. Proposed Mobile Cloud Methodology

Mobile Cloud Computing bases offered by VMware and open MPI innovation can be utilized to meet the necessities effectively [34]. C and C++ languages can be utilized to program the issue and execute the design of High-Performance Mobile Cloud Computing for proposed stages. Consolidated performance of VM machines is altogether higher than single equipment machine/processor, because of its high asset capacities. The devices given by versatile distributed computing, for example, VMware alongside open MPI, to run an issue in a parallel design can upgrade the execution of use on Mobile Cloud Figuring Infrastructure [35].

High computational tasks have been accomplished through hardware solution which is costly. The proposed solution is software based, one which comprises virtual machines having a hypervisor installed in it and programming utilities/software to execute the required computations. This software is installed on the VMware machines to run the distributed computational task.

The function "double calcBondStrechTerms (int begin, int finished)" takes two integer variables "start" and "end" as input and calculates results as a variable type double. The required parameters and constants are obtained from a file named "amber parameters." It uses the parameters mentioned in Table 2.

The function is implemented and mentioned as "double calcAngleBendTerms (int begin, int finished)" in our code. This function takes the two integer variables "start" and "end" as input and calculates results as variable type double. Another function "calcDHAngleterms (int begin, int finished)" is mentioned as "double type." This function takes the two integer variables "start" and "end" as input and calculates results as a variable type double. The required parameters and constants are obtained from a file named "amber parameters." It uses the parameters mentioned in Table 3 from the parameters required to calculate the dihedral angle terms and half nonbonded terms as mentioned in Table 3.

The function "double calcHalfNBterms (int begin, int finished)" takes the two integer variables "start" and "end" as input and calculates results as a variable type double. The function named as "double calcNBterms (int begin, int finished)" takes the two integer variables "begin" and "finished" as input and calculates results as a variable type double. Another function that is mentioned as "double calcHBterms (int begin, int finished)" in our code takes the two integer variables "begin" and "finished" as input and calculates results

Table 3: Parameters required to calculate dihedral angle terms and half nonbonded terms.

Sr. No.	Parameter	Description
(1)	"dihedralAngleTerms [ix7]"	All the four values of dihedral angle terms according to functional form
(2)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array
(3)	"phiDeg"	Bend of angle represented in degree keeping in view radian to degree relationship
(4)	"app"	Intermediate variable to calculate phiDeg
(5)	"energy"	The partial value of energy due to dihedral angle terms is calculated and stored in it
(6)	"coulombFactor"	It is calculated using permittivity and permittivity scale and a constant value 116
(7)	"halfNonBondedTerms [ix4]"	All the four values of half nonbonded terms according to functional form
(8)	"charge$i$"	Value of charge on atom i
(9)	"charge$j$"	Value of charge on atom j
(10)	"coordinates [atomix3]"	X, Y, Z coordinates of atom stored in array

Table 4: Names and IPs of experimental machines.

Sr. No.	Machine	Hardware distribution	Name	IP
(01)	VM1	2 GB RAM, 2 processor, 40 GB HD	Server	192.168.2.50
(02)	VM2	1 GB RAM, 1 processor, 40 GB HD	Client 1	192.168.2.51
(03)	VM3	1 GB RAM, 1 processor, 40 GB HD	Client 2	192.168.2.52
(04)	VM4	1 GB RAM, 1 processor, 40 GB HD	Client 3	192.168.2.53
(05)	VM5	1 GB RAM, 1 processor, 40 GB HD	Client 4	192.168.2.54

as a variable type double. Both functions required parameters and constants obtained from a file named "zamberparameters." This program can run as multiple processes on different cores of coprocessor and processor.

5. Experimental Equipment and Setup

Our experimental setup mainly up consists of the following hardware and software: the 2x laptops having Core i5 and Core i7 with VMware Workstation software and 1x WiFi network connection as shown in Figure 4. All these systems have Linux OPEN SUSE as the operating system on machines with NetBeans Development toolkit for the AMBER calculations. Experimental Setup details are given in Table 6: In the proposed solution, there are two laptops Core i5 and Core i7. There are five virtual machines (VMs) on two laptops and details are shown in Table 4. The hardware details of the machines are given in Table 6. The hardware and associated software are mentioned for the user understanding.

The server is the central machine that distributes multiple tasks to various clients through MPI.

Open MPI is the core resource to distribute multiple tasks to various clients attached to the server of the virtual network. All machines have their SSH keys shared among one another to achieve passwordless connection to each machine for calculations. VMware assigns hardware resources to machines. NetBeans is installed on server for program compilation and execution. The screen of the server distributed computational process over the computing nodes as shown in Figure 5. Linux Open SUSE OS is installed on each machine as a hypervisor. All machines are on the same network, that is, 192.168.2.0/42. VMs are connected through a virtual switch; in response, all clients send their tasks results to the main server. The server then calculates the Grande total result received from all clients. Computation node with completion time, CPU, PID, and memory is shown in Figure 5.

6. Results and Discussion of HPCC

High-performance cloud computing is an advanced technique that relies on virtualization and parallelism. To simulate this technique, multiple VMs are created in the simulated cloud. Resources are assigned to these VMs. One of the machines is the core machine named as a server. In Table 5, the comparison is being made among various solutions for molecular dynamics calculations. The comparison is carried out by HPC with Biomer. Biomer is software that runs on the single machine and lacks parallel mechanism like MPI. Hybrid parallel computing is a hybrid parallel technique that uses multiple hardware processing cores along Intel MPI to reduce the time for molecular dynamic tasks calculation. It splits and distributes tasks to other submachines named as clients. Clients in response send their result to the server. Table 5 shows that HPCC has much lower time as compared to the previous method.

The experiment consists of multiple numbers of processes like 1, 2, 12, 57, and 58. These numbers have been taken from HPC for apple to apple comparison. It also shows that, as none

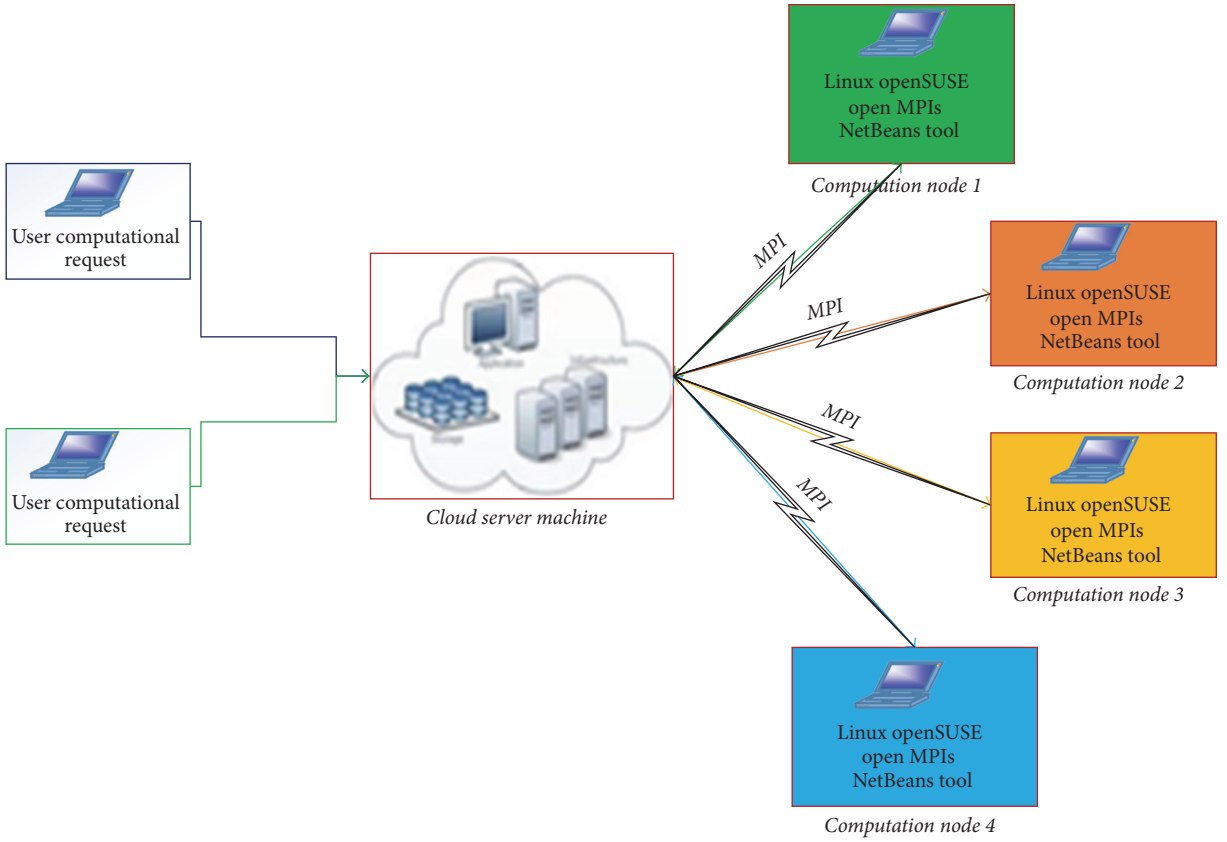


Figure 4: High-performance cloud computing simulation model.

Table 5: Comparison of processing time required for calculation of energy-based tasks.

No. of processes	HPC (second)	HPCC (second)	Biomer CPU (second)
1	N/A	N/A	5073
2	676	153	N/A
12	78	25	N/A
57	23	7	N/A
58	23	5	N/A

Table 6: Hardware and software comparison of HPC and HPCC.

Sr. No.	Attributes	Hybrid parallel computing (HPC)	High-performance cloud computing (HPCC)
(1)	Name	Intel Xeon Phi™	2x Dell laptop
(2)	Processor	Intel3120A	Core i7, i5
(3)	RAM	6 GB Ram	8 GB
(4)	Clock speed	1.00 GHz	2.7 GHz~3.5 GHz
(5)	Cores	57 Cores	No extra hardware
(6)	Card/VM	2x 2.1 GHz CPU, 2 GB RAM	4x VMs
(7)	VM specific	No VM	3x (1 processor, 1 GB RAM) 1x (1 processor, 1 GB)
(8)	Power supply	Additional required	No additional
(9)	Watt	400 for Phi TM Card	0
(10)	OS	Cent OS	Windows 7 & 10
(11)	Kernel type	Linux 2.6.32-431	SUSE Linux
(12)	Libraries	IMPI	MPI
(13)	Version	Intel Cluster Studio XE 2013	Server machine has NetBeans
(14)	Tested	Simple, compiled using IMPI, running on the coprocessor	Simple, compiled using MPI, Intel CPU on SUSE Linux

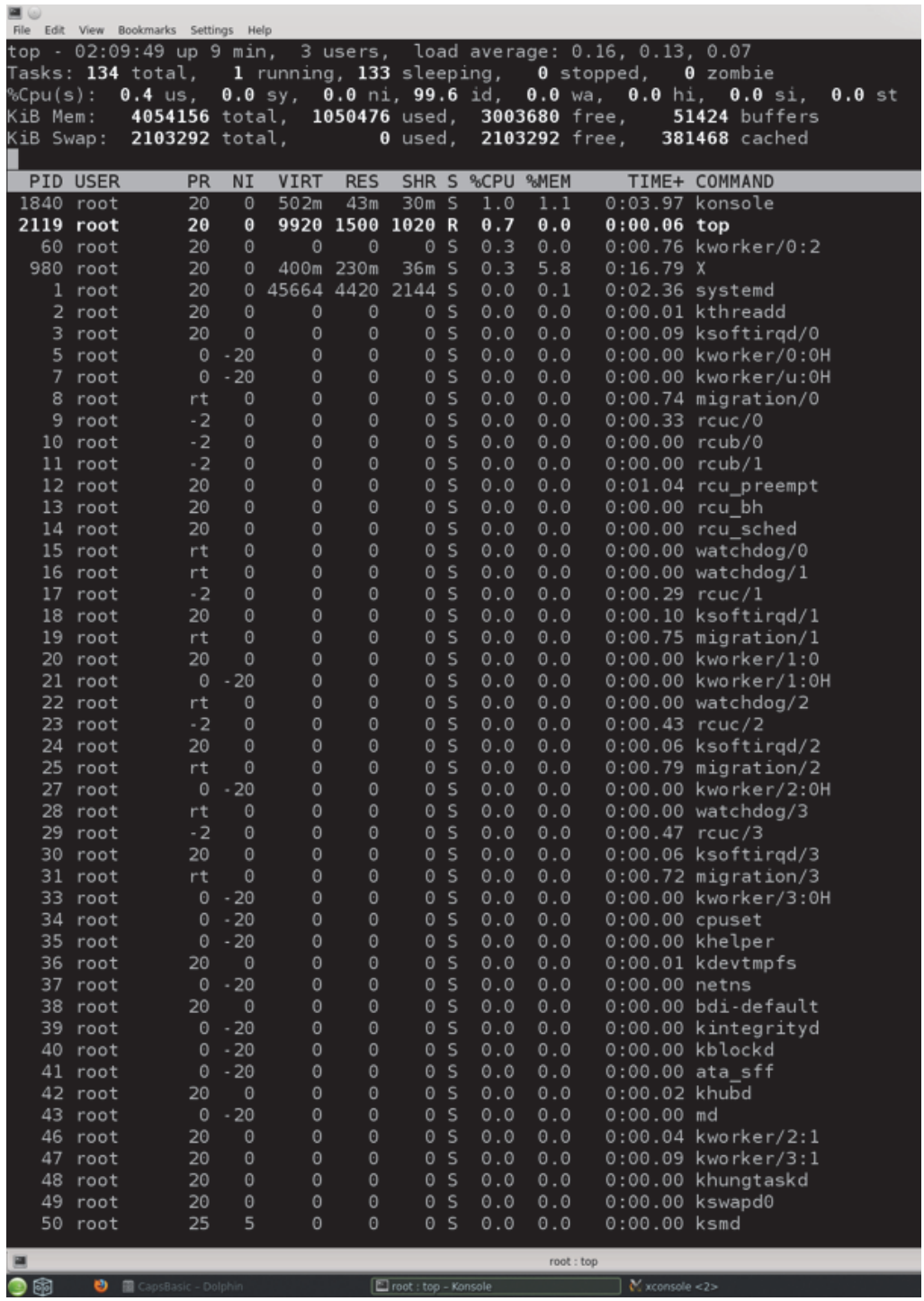


Figure 5: Computation node with completion time, CPU, PID, and memory.

of processes increases, time for calculating tasks reduces just because of parallel techniques as shown in Figure 6.

In Table 6, there is the comparison of hardware resources used in HPCC and HPC solution. Hybrid parallel computing (HPC) is much resource consuming as compared to high-performance cloud computing which is based on virtualization. HPC hardware is much costly as compared to HPCC.

Table 6 shows that HPCC mostly relies on software as compared to HPC which relies on costly hardware. The processing time has been obtained by executing the same energy calculation, and complexity of Biomer Java-based software is 5073 sec which is reduced to hundreds of seconds by using cloud computing in combination with parallel computing. It is because Biomer software is Java-based and

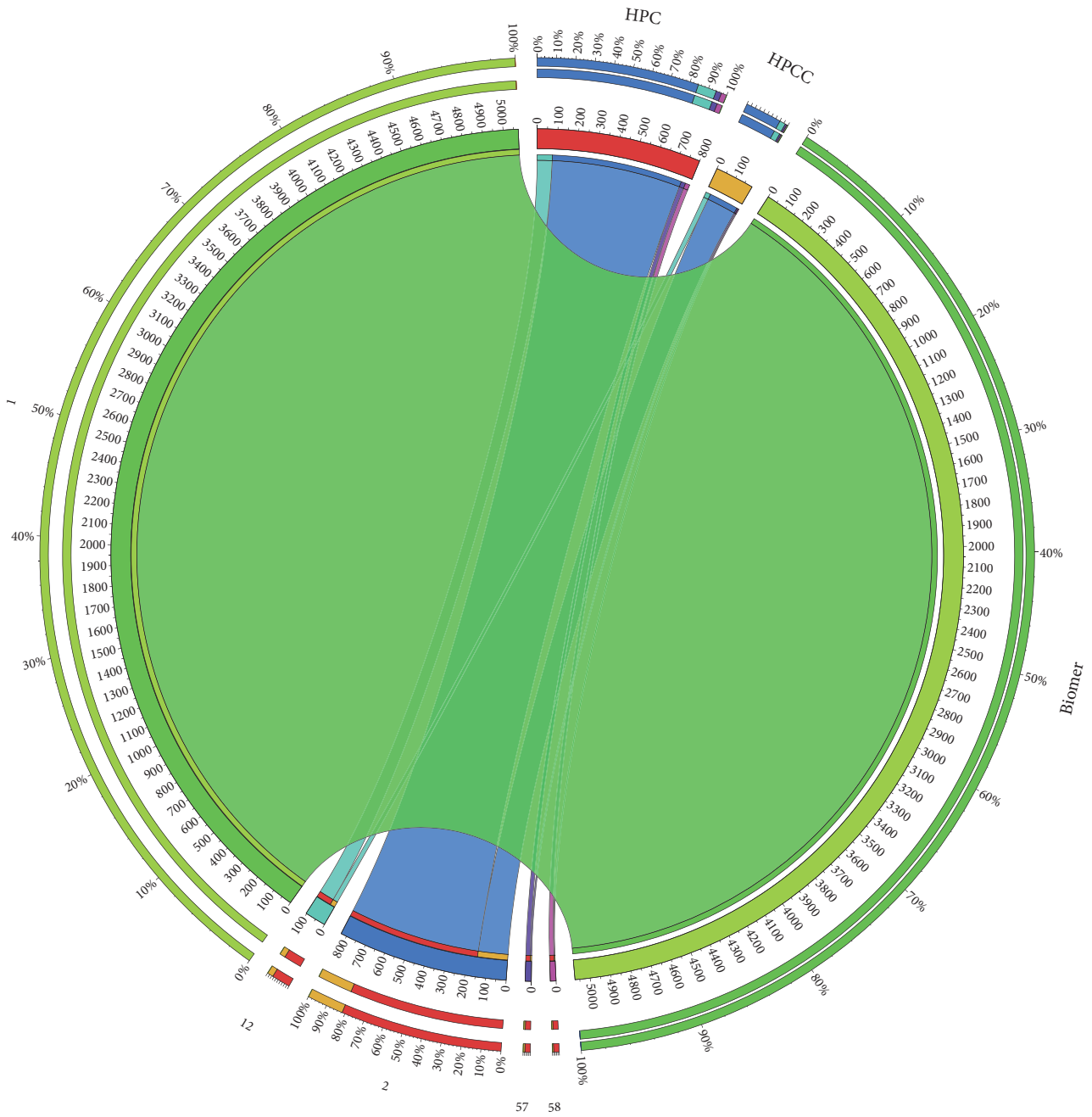


Figure 6: Comparison of HPC and HPCC time required for calculation of energy.

runs as a single process without using the mechanism of the parallelism. It has further to reduce calculation time by increasing number of process and machines resources. Figure 7 shows the comparative result of HPCC and HPC techniques.

It would get more efficient result by increasing number of the process to be run using parallelism and using high-end desktop machines and optimized network media. In another comparison depicted in Figure 8, when number of processes has been increased to 100% of previous value, calculation time is reduced to 50% of previous value. Values for number of processes are 10, 20, 40, and 80, processing time is reduced to one-half of previous value, and efficiency is increased to 100% as shown by the graph in Figure 8.

If X is the number of processes to be selected for calculation and y is the calculation time for number of processes of the previous value, then X will be calculated through the equation $X = (1/2)Y$. Figure 8 shows that when number of process increased up to double of the previous value, its calculation time is reduced up to 1/2 of the previous value of calculation time. The result is verified in Table 5. In the table, there is the comparison of software resources used in HPCC and HPC solution. HPCC uses VMware Workstation as virtualization and cloud computing tool. Physical machine

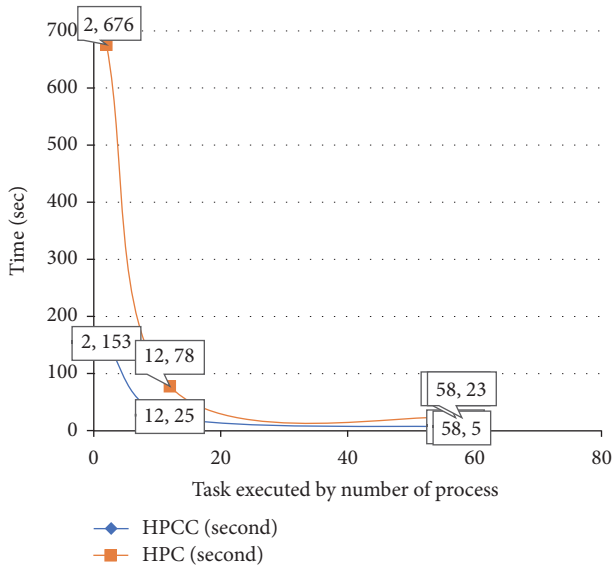


Figure 7: Comparison chart of HPCC and HPC processes.

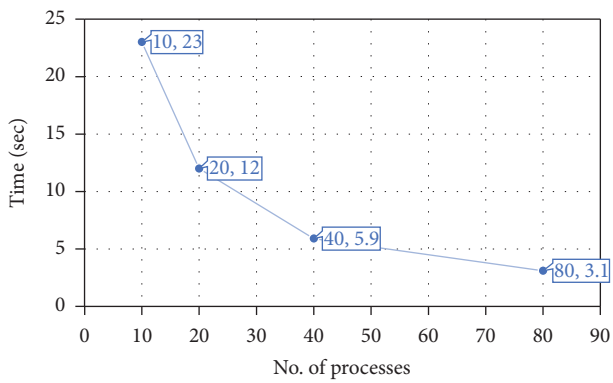


Figure 8: Comparison of HPCC in terms of number of processes versus execution time.

has Windows OS and VMs have open SUSE Linux flavour. Open MPI is used for the parallel tool. HPCC mostly relies on software as compared to HPC which relies on costly hardware.

7. Conclusion and Future Work

It is concluded from this research that big data in the field of molecular dynamics is a high computational task and requires large computational resources and advanced techniques to meet the requirements. Data can be extracted, processed, and analyzed in AMBER required cloud computing techniques which are suitable for the parallel execution of tasks. It provides flexibility and scalability of resources for calculation of heavy computational tasks. It is a homogeneous mobile cloud platform consisting of laptops and virtual machines as processing nodes in combination with dynamic parallelism. Multiple numbers of processes have been chosen at run time to distribute the task among the computing nodes. Message Passing Interface is used to attain task-based and data-based parallelism. It used shared hardware resources for VMs for achieving better results. In future high-end mobile and smartphone using fast and efficient networking media can be used for better results. MPI can further be tuned and programmed for dynamic parallelism. MPI can also be integrated with another Linux flavours as well. Moreover, it is also implementable for the big data for the enterprises and analytics.

Conflicts of Interest

All the authors confirm that there are no conflicts of interest regarding the publication of this research article.

Acknowledgments

This research is supported by the Department of Computer Science, University of Engineering and Technology, Taxila.

References

[1] Y. Saleem, M. Munwar Iqbal, M. Amjad et al., "High security and privacy in cloud computing paradigm through single sign on," *Life Science Journal*, vol. 9, no. 4, pp. 627–636, 2012.

[2] M. H. R. Pereira, C. L. de Souza, F. L. C. Pádua, G. D. Silva, G. T. de Assis, and A. C. M. Pereira, "SAPTE: A multimedia information system to support the discourse analysis and information retrieval of television programs," *Multimedia Tools and Applications*, vol. 74, no. 23, pp. 10923–10963, 2015.

[3] C. Yang, Y. Xu, and D. Nebert, "Redefining the possibility of digital Earth and geosciences with spatial cloud computing," *International Journal of Digital Earth*, vol. 6, no. 4, pp. 297–312, 2013.

[4] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: from big data to big impact," *MIS Quarterly: Management Information Systems*, vol. 36, no. 4, pp. 1165–1188, 2012.

[5] M. M. Iqbal, Y. Saleem, K. Naseer, and M. Kim, "Multimedia based student-teacher smart interaction framework using multi-agents in eLearning," *Multimedia Tools and Applications*, pp. 1–24, 2017.

[6] B. Joó, D. D. Kalamkar, K. Vaidyanathan et al., "Lattice QCD on Intel® Xeon Phi™ coprocessors," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 7905, pp. 40–54, 2013.

[7] A. Gupta, A. Mehrotra, and P. M. Khan, "Challenges of cloud computing & big data analytics," in *Proceedings of the 2nd International Conference on Computing for Sustainable Global Development, INDIACom 2015*, pp. 1112–1115, ind, March 2015.

[8] P. Malik, "Governing Big Data: Principles and practices," *IBM Journal of Research and Development*, vol. 57, no. 3/4, pp. 1:1–1:13, 2013.

[9] C. J. Newburn, S. Dmitriev, R. Narayanaswamy et al., "Offload compiler runtime for the intel® Xeon Phi coprocessor," in *Proceedings of the 2013 IEEE 27th International Parallel and Distributed Processing Symposium Workshops and PhD Forum, IPDPSW 2013*, pp. 1213–1224, usa, May 2013.

[10] S. Ahmad, S. Xu, B. Li, and S. Ahmad, "Research and implementation of Hybrid Parallel Computing for force field calculation,"

in *in Electronics, Information Technology and Intellectualization: Proceedings of the International Conference EITI 2014*, Shenzhen, China, 2015.

[11] J. C. A. Boeyens, "A molecular-structure hypothesis," *International Journal of Molecular Sciences*, vol. 11, no. 11, pp. 4267–4284, 2010.

[12] W. D. Cornell, P. Cieplak, C. I. Bayly et al., "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, pp. 5179–5197, 1995.

[13] G. M. Tihfon, S. Park, J. Kim, and Y.-M. Kim, "An efficient multi-task PaaS cloud infrastructure based on docker and AWS ECS for application deployment," *Cluster Computing*, vol. 19, no. 3, pp. 1585–1597, 2016.

[14] G. D. M. Seabra, R. C. Walker, M. Elstner, D. A. Case, and A. E. Roitberg, "Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package," *The Journal of Physical Chemistry A*, vol. 111, no. 26, pp. 5655–5664, 2007.

[15] M. Allen and D. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, UK, 1987.

[16] V. Gapsys, D. Seeliger, and B. L. De Groot, "New soft-core potential function for molecular dynamics based alchemical free energy calculations," *Journal of Chemical Theory and Computation*, vol. 8, no. 7, pp. 2373–2382, 2012.

[17] S. Mittal, "A survey of techniques for architecting and managing asymmetric multicore processors," *ACM Computing Surveys*, vol. 48, no. 3, article no. 45, 2016.

[18] B. D. Veerasamy and G. Nasira, "Exploring the Contrast on GPGPU Computing through CUDA and OpenCL," *i-manager's Journal on Software Engineering*, vol. 9, no. 1, pp. 1–8, 2014.

[19] H. Zheng, K. M. Langner, G. P. Shields et al., "Data mining of iron(II) and iron(III) bond-valence parameters, and their relevance for macromolecular crystallography," *Acta Crystallographica Section D: Structural Biology*, vol. 73, no. 4, pp. 316–325, 2017.

[20] L. F. L. Oliveira, N. Tarrat, J. Cuny et al., "Benchmarking Density Functional Based Tight-Binding for Silver and Gold Materials: From Small Clusters to Bulk," *The Journal of Physical Chemistry A*, vol. 120, no. 42, pp. 8469–8483, 2016.

[21] A. Verma, O. Uzun, Y. Hu et al., "Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles," *Nature Materials*, vol. 7, no. 7, pp. 588–595, 2008.

[22] Molecular_dynamics, ed: Wikipedia, 2013.

[23] G. H. Fredrickson, V. Ganesan, and F. Drolet, "Field-theoretic computer simulation methods for polymers and complex fluids," *Macromolecules* , vol. 35, no. 1, pp. 16–39, 2002.

[24] C. Delgado, S. Y. Stevens, N. Shin, and J. Krajcik, "A middle school instructional unit for size and scale contextualized in nanotechnology," *Nanotechnology Reviews*, vol. 4, no. 1, pp. 51–69, 2015.

[25] Y. Duan, C. Wu, S. Chowdhury et al., "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations," *Journal of Computational Chemistry*, vol. 24, no. 16, pp. 1999–2012, 2003.

[26] F. Magoulès, M. Parsons, and L. Smith, *Innovative Algorithms for Extreme Scale Computing*, SAGE Publications Sage, London, England, 2015.

[27] J. Vetter, A. Almgren, P. DeMar, K. Riley, K. Antypas, D. Bard et al., *Advanced Scientific Computing Research Exascale Requirements Review. An Office of Science review sponsored by Advanced Scientific Computing Research, September 27-29, 2016, Rockville, Maryland*, Argonne National Lab. (ANL), Argonne, IL, USA, 2016, Argonne Leadership Computing Facility 2017.

[28] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *Proteins: Structure, Function, and Bioinformatics*, vol. 65, no. 3, pp. 712–725, 2006.

[29] J.-s. Shyr, *Adaptive scheduling of function cells in dynamic reconfigurable logic*, Google Patents, 2005.

[30] R. A. Bridges, N. Imam, and T. M. Mintz, "Understanding GPU power: A survey of profiling, modeling, and simulation methods," *ACM Computing Surveys*, vol. 49, no. 3, article no. 41, 2016.

[31] D. Barr, A. Basden, N. Dipper, and N. Schwartz, "Reducing adaptive optics latency using Xeon Phi many-core processors," *Monthly Notices of the Royal Astronomical Society*, vol. 453, no. 3, pp. 3222–3233, 2015.

[32] E. Wang, Q. Zhang, B. Shen et al., *High-Performance Computing on the Intel Xeon Phi*, Springer International Publishing, 2014.

[33] J. Kim, K. Jang, K. Lee, S. Ma, J. Shim, and S. Moon, "NBA (network balancing act): A high-performance packet processing framework for heterogeneous processors," in *Proceedings of the 10th European Conference on Computer Systems, EuroSys 2015*, France, April 2015.

[34] G. Neiger, A. Santoni, F. Leung, D. Rodgers, and R. Uhlig, "Intel Virtualization Technology: Hardware Support for Efficient Processor Virtualization," *Intel Technology Journal*, vol. 10, 2006.

[35] P. Raj, A. Raman, D. Nagaraj, and S. Duggirala, "High-Performance Grids and Clusters," in *High-Performance Big-Data Analytics*, pp. 275–315, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8261746, 11 pages
https://doi.org/10.1155/2018/8261746

Research Article

MQTT Security: A Novel Fuzzing Approach

Santiago Hernández Ramos,[1] **M. Teresa Villalba**,[2] **and Raquel Lacuesta**[3]

[1]*Telefónica Digital, Madrid, Spain*
[2]*Universidad Europea de Madrid, Madrid, Spain*
[3]*Universidad de Zaragoza, Teruel, Spain*

Correspondence should be addressed to Santiago Hernández Ramos; santiago.hernandezramos@telefonica.com

Received 30 September 2017; Accepted 10 January 2018; Published 26 February 2018

Academic Editor: Syed H. Ahmed

The Internet of Things is a concept that is increasingly present in our lives. The emergence of intelligent devices has led to a paradigm shift in the way technology interacts with the environment, leading society to a smarter planet. Consequently, new advanced telemetry approaches appear to connect all kinds of devices with each other, with companies, or with other networks, such as the Internet. On the road to an increasingly interconnected world, where critical devices rely on communication networks to provide an essential service, there arises the need to ensure the security and reliability of these protocols and applications. In this paper, we discuss a security-based approach for MQTT (Message Queue Telemetry Transport), which stands out as a very lightweight and widely used messaging and information exchange protocol for IoT (Internet of Things) devices throughout the world. To that end, we propose the creation of a framework that allows for performing a novel, template-based fuzzing technique on the MQTT protocol. The first experimental results showed that performance of the fuzzing technique presented here makes it a good candidate for use in network architectures with low processing power sensors, such as Smart Cities. In addition, the use of this fuzzer in widely used applications that implement MQTT has led to the discovery of several new security flaws not hitherto reported, demonstrating its usefulness as a tool for finding security vulnerabilities.

1. Introduction

Today, cities face complex challenges, including sustainable urban development, reduction of pollution and energy consumption, and safety [1]. IoT (Internet of Things) is considered the core technology for building Smart Cities, as it is based on the concept "everything can be connected to the Internet." The development of cheaper sensors and other devices, as well as the adoption of cloud services, is providing new opportunities to develop new services for improving quality of life in cities. As cities grow, interest in exploring new IoT technologies increases. Some examples of how IoT technologies can support the building of Smart Cities are as follows:

(i) Smart street lights with sensors for detecting cars' movement and sending data to control when to switch them on or off to save energy

(ii) Reducing water consumption in parks

(iii) Health personnel attending to citizens in emergency situations with access to their medical records in real time [2–4]

(iv) The consumption and regulation of electricity controlled by smart meters and sensors that send data in real time [3, 5, 6]

Due to IoT, technology-enabled devices located in different places communicating with each other and generating large volumes of data, information becomes more difficult to protect [7]. Compromised availability, integrity, or confidentiality of these data can have an adverse and direct effect on people's lives [8]. Consequently, there is a need to implement mechanisms that verify the security of IoT devices, the network protocol they use to exchange information, and the applications developed for them.

In this paper, we propose a framework to improve the security of applications implementing the protocol MQTT

(Message Queue Telemetry Transport), a widely used protocol for sharing data exchanged between IoT devices. MQTT is an extremely simple and lightweight messaging protocol, with a publish/subscribe architecture, designed to be straightforward to deploy, and capable of supporting thousands of clients with a single server. In addition, MQTT provides reliability and efficiency in adverse conditions. All these features make this protocol one of the most used protocols for the communication between smart devices, with a high number of applications based on it, increasing rapidly over time [9, 10].

To test the security of the applications that implement MQTT, we have created a framework based on a verification technique called fuzzing. Fuzzing is a testing technique for finding vulnerabilities in software applications [11] by sending unexpected input data to target systems and then monitoring the results. Typically, it consists of an automatic or semiautomatic process, which comprises the sending and repeated manipulation of data to the system under study. All fuzzers can be classified into two broad categories [12]: mutation-based and generation-based fuzzers. Mutation-based fuzzers apply mutations on existing data samples to create the test space, while generation-based fuzzers create test cases from scratch by modelling the target protocol or file format. As generation-based approaches are more complex and time-consuming, we focus on mutation-based fuzzer approaches along with a novel fuzzing technique based on templates. The aim of this technique is to reduce the effort and increase the productivity of users when performing security verification of the applications that implement the MQTT protocol. This new technique allows completely automated generation of a template with the fields we want to test for each network packet. It also enables definition of other specifications, such as the fields for which we want to filter the traffic or the values that we want to insert or send by default.

The rest of the paper is organized as follows. In Section 2, we briefly explain the basic concepts needed to understand the presented work. Section 3 discusses the related work regarding MQTT protocol security and the modern fuzzing approaches. Section 4 deals with the basic elements of the framework and the implementation of the concepts and methods discussed above. Section 5 then introduces the architecture of the fuzzer, and Section 6 shows the results of the experimentation phase. Finally, the conclusions are presented.

2. Background and Motivation

2.1. Message Queue Telemetry Transport. MQTT uses a publish/subscribe messaging pattern that enables a loose coupling between the information provider, called the *publisher*, and consumers of the information, called *subscribers*. This is achieved by introducing a message broker between the publishers and the subscribers.

Compared with the traditional point-to-point pattern, the advantage of this model is that the publishing device or the application does not need to know anything about the subscribing one, and vice versa. We can distinguish three MQTT essential concepts that will remain present throughout the development of the paper.

Table 1: MQTT fixed header.

Bit->	7	6	5	4	3	2	1	0
Byte 1	Msg type				DUP	QoS		Retain
Byte 2	Remaining length							

Table 2: Some solutions that use MQTT.

Brokers	Clients	Smart home
Mosquitto	CocoaMQTT	Homegear
ActiveMQ	emqttc	Domoticz
hbmqtt	mqtt-client	Lelylan
HiveMQ	M2Mqtt	cul2mqtt
Moquette	mqtt_cpp	aqara-mqtt
Mosca	mqttex	Home.Pi
VerneMQ	Paho	Home Assistant
hrotti	rumqtt	pimatic
SurgeMQ	hbmqtt	FHEM

(i) Topics. The *publishers* are responsible for cataloguing the messages they send in topics. A topic defines the content of a message or a category in which the message can be classified. Topics are important because while in the point-to-point protocols messages are sent to a specific address, in a publish/subscribe pattern, messages are distributed based on the selected topics by the subscriber. By subscribing to a particular topic, the subscriber will receive all messages sent with that topic by any publisher.

(ii) Client. MQTT clients connect to a broker to exchange messages. They must subscribe to topics and can publish information to other entities connected to the same broker by providing a topic.

(iii) Broker. MQTT brokers are servers acting as intermediaries for the messages. MQTT protocol messages' format consists of three parts: a fixed header, shown in Table 1; a variable header; and a payload. Fuzzers consider the fields and positions of the header for inserting data to perform the fuzzing process.

MQTT is one of the most used protocols worldwide as shown in Table 2.

2.2. Fuzzing Processes. The phases of a fuzzing process are highly variable and depend on many factors, such as the application being tested or the programmer's experience [13]. However, there is a set of basic steps that are always followed, regardless of the approach or application being analysed. In the development of the tool that will be presented below, the following phases have been considered:

(i) Identifying an objective: the first step in every fuzzing process consists of identifying the target which could be an application, a protocol, or even the function of a specific library. The target here is both the MQTT protocol and the applications implementing the protocol.

(ii) Identifying the entry points: almost all exploitable vulnerabilities are caused by applications that accept user values processed without being properly checked in advance. Enumerating the input vectors is one of the crucial aspects for the fuzzing process to succeed. In the end, anything that can be sent from the client to the target system should be considered as an input vector. This includes headers, filenames, environment variables, and registry keys.

(iii) Generating the fuzzing data: once the input vector has been identified, we must generate appropriate data to perform the fuzzing process. A high degree of automation generating the test cases is important, as numerous cases must be generated.

(iv) Executing the test cases: this step is closely linked to the previous one and consists of the process of sending the data packets to the target system. As in the previous stage, process automation is essential.

(v) Exception monitoring: a vital part of the fuzzing process is monitoring exceptions. Causing the crash of the target system after sending numerous data packets has no benefit if the particular packet that caused the error cannot be determined. Monitoring can take many forms and is closely linked to the target system and the type of fuzzing being performed.

3. Related Work

3.1. Security in MQTT Protocol. Although research on MQTT security is still scant, some incipient work has been presented about its security issues. Almost all security problems that arise are related to the state in which the protocol works by default. Because MQTT is a simple protocol designed for devices with low processing power, by default, the protocol tries to minimize the processing needed to exchange messages, which means that serious security problems arise. Most of these shortcomings can be solved with an adequate protocol configuration. The following are some of the most common security issues that can be solved through proper protocol configuration [14, 15]:

(i) Lack of authentication: the MQTT protocol does not provide a secure authentication mechanism by default, which can lead to spoofing the identity of some of the participants in the communication or the sending of unauthorized data. This problem can be easily solved by configuring the protocol features adequately. When it comes to authentication, the protocol itself provides username and password fields in the CONNECT message enabling clients to send a username and password when connecting to an MQTT broker.

(ii) Lack of authorization: MQTT clients, after connecting to a broker, can publish messages or subscribe to topics. Each authenticated client can publish and subscribe to all kinds of topics even without proper authorization. This may be a significant problem, because the protocol itself does not provide any mechanism to carry it out and therefore the responsibility lies with the broker. In spite of this, it can be easily solved through the implementation of topic permissions on the broker side [16].

(iii) Lack of confidentiality: MQTT relies on TCP as transport protocol, which means that by default the connection does not use an encrypted communication. This means that packets can be spied on by an attacker listening on the same network. To avoid this, almost all MQTT brokers allow encryption of the whole MQTT communication, using TLS instead of plain TCP.

(iv) Lack of integrity: when MQTT systems have untrusted clients or unidentified MQTT clients have access to the MQTT broker and topics, data integrity of sent messages should be checked, especially when TLS is not used. MQTT supports three mechanisms to provide integrity to exchanged packets: Checksum, MAC, and Digital Signatures.

Other approaches have been also presented by some authors. Moreover, some research has tried to deal with the general problems of IP-based protocols used by IoT devices, one of which is MQTT. In these cases, the authors focus on the security of this type of device as part of a broader spectrum, treating the layers of protection that can wrap around the TCP/IP protocol and the security architectures and models that best fit IoT networks [17]. In addition to confidentiality, other security features have been addressed; [18] deals with the problem that smart devices have when they do not have enough processing capacity to use asymmetric encryption algorithms to perform authentication tasks, and it proposes a new authentication approach based on operations that consume few resources, such as hash functions or OR operations. Some other interesting approaches focus on how to force compliance with the optional security features that the MQTT protocol can implement. *SecKit* is a model-based security toolkit that tries to force the use of a series of security policies, so that the protocol implements some protection measures that are not found in its default implementation [19]. There is also research that continues to focus on the security limitations that this protocol poses by design and proposes frameworks to improve their security in the transporting of information between the parties involved in the connection by adding extra layers such as SSL/TLS [20, 21].

The considerations that appear at the beginning of this section show that the protocol security flaws are related to its operation and, in particular, to the way in which it exchanges information. This paper aims to contribute to the improvement of the security of the MQTT protocol regarding the verification of the devices that implement it. When applications that implement the MQTT protocol process a package incorrectly, serious security failures such as denial of service or remote execution of arbitrary code can occur [22, 23], even when the security measures mentioned above are met. Evaluating applications that implement such a protocol in the different parts of the connection (client and broker) to verify their behaviour when receiving incorrect or unexpected data can help to avoid certain serious security breaches.

3.2. Modern Fuzzing. As has been said in previous sections, there are many types of fuzzing and many ways to perform this technique, including IoT fuzzing as the one presented for the Modbus protocol [24] by the *IoT Systems book* [25]. However, in general, as Aitel stated in his paper [26], modern fuzzing tries to solve three major problems with respect to traditional fuzzing.

(i) If the network protocol is defined by an API with which the client and server are implemented, it is very likely that these predefined functions will make certain checks on the data that is sent and will consequently have an indirect influence on the fuzzing process.

(ii) Even given complete knowledge of the protocol, creating a client for a protocol can be a considerable undertaking, and that client is rarely portable to other protocols, even those of a similar nature.

(iii) Often, testers have only limited knowledge of the protocols under attack or of the ways the protocol may break.

To solve this, modern fuzzing tools, like Boofuzz [27], SNOOZE [28], and KiF [29] or [30], among others, propose a block-based approach, which consists of decomposing the protocols into length fields and data fields and providing the user with a framework for creating such tools without having to worry about the control fields (such as lengths or checksums) of the lower layers. This is the approach that most current frameworks use, and it has a very simple foundation. If, when a network packet consisting of several layers is available, with the upper layers being the application ones and the lower layers being the physical layers, we would like to perform fuzzing testing on one of the application layers, it would not be enough to enter the testing value and send the message, since the underlying layers may contain control fields, which if not updated correctly would lead to rejection of the packet upon reaching the server, before the value inserted was processed. To solve this problem, some structures called blocks were proposed. In them, a series of variables is grouped previously defined by the framework, which occupy a specified size. A set of these variables form a block, and the blocks can be opened and closed as follows [31]:

```
s_block_size_binary_bigendian_word("somepacketdata");
s_block_start("somepacketdata");
s_binary("01020304");
s_block_end("somepacketdata");
```

What is achieved with this is that the control fields of the lower layers are recalculated automatically, once all the blocks have been closed and, consequently, the user does not have to worry about processing them.

If we focus on the particular topic of fuzzing the MQTT protocol, very few references or tools can be found about it. The only public tool of which the authors have evidence is mqtt-fuzz [32], whose main utility is to verify the protocol in a fast and traditional way, without providing too much complexity. In addition, other methods of fuzzing or formal testing have been presented for the MQTT protocol, albeit with a different aim. This is the case with [33], in which the authors discuss formal methods of network protocol verification through finite-state machines and labelled transition systems. The focus is on demonstrating how most of the implementations of MQTT do not meet the standard. CG-Fuzzing is a fuzzy algorithm for ZigBee, with a focus on generating an efficient number of test cases.

3.3. Proxy Fuzzing. Proxy fuzzing is a widespread and a barely studied technique, resulting from some of its current limitations. Some work has been carried out in relation to this technique, such as ZAP Proxy [34], Burp Proxy [35], ProxyFuzz [36]. What all this work has in common is that the fuzzer must be placed in the middle of the connection, between the client and the server, to serve as a relay agent. To effectively accomplish this task, both the client and the server must be configured manually or automatically by some IP trickery, for example, ARP spoofing. This allows the client and the server to look for one another at the address of the proxy, so the client sees the proxy as the server, and vice versa. This fuzzing method provides several improvements over previous processes, such as simplicity of use. However, it is a difficult technique to implement, which is why the tools for implementing this technique are barely known and introduce extremely basic fuzzing techniques. This technique has led to a patent [37].

4. Methods

In this section, we will show how we have implemented the methods mentioned in the previous sections, along with other new approaches to creating a fuzzing tool for the MQTT protocol.

4.1. Fuzzing MQTT Messages. The process of fuzzing a protocol or the applications that implement it entail knowing in some way the specification of said protocol, either through its public documentation or by reverse engineering techniques. Once we know its specification and we can interpret the bytes of a package, we must select the packages and fields that are of interest for inserting information with the intention of verifying that the application that processes them does so correctly.

In the case of MQTT, no reverse engineering process is required, since its specification is public [38]. Therefore, we need only look in the specification documents for the type of

Table 3: Types of MQTT messages.

Packet	Description
CONNECT	Connect to the server
CONNACK	Ack of connect msg
PUBLISH	Publish a topic
PUBACK	Ack of publish msg
PUBREC	Publication received
PUBREL	Publication sent
PUBCOMP	Publication completed
SUBSCRIBE	Client subscription
SUBACK	Ack of subscribe msg
UNSUBSCRIBE	Unsubscribe petition
UNSUBACK	Ack of unsubscribe msg

Table 4: Publish packet variable header nonnormative example.

Byte position	Description
Topic name	
Byte 1	Length MSB (0)
Byte 2	Length LSB (3)
Byte 3	"a" (0x61)
Byte 4	"/" (0x2F)
Byte 5	"b" (0x62)
Packet identifier	
Byte 6	Packet identifier MSB (0)
Byte 7	Packet identifier LSB (10)

packages which are exchanged and the fields that are in their variable header and payload (Table 3).

If we look a little more in depth at the type of packets exchanged by the protocol, we quickly realize that the message *PUBLISH* is likely to be the one in which most information is transmitted and therefore the one in which the most processing is performed by the applications that implement the protocol. Once we have identified this type of packages (PUBLISH, CONNECT, SUBSCRIBE, etc.), we study their variable header (Table 4) to select the type of fields and the fields' positions in bytes, into which the test cases will be inserted to carry out the fuzzing process. Once the fields where the test cases are inserted have been selected, we look for the control fields, which will be recalculated once the test case is inserted. Finally, we look for a field that unequivocally identifies the package in order to be able to filter it "on the fly."

4.2. Advance Proxy Fuzzing. To apply the fuzzing process, we use the fuzzing proxy technique explained in previous sections.

As we have already stated, this technique is not very widespread, and the tools that perform it are outmoded and present a great degree of deficiency compared to modern techniques. However, if we study in depth the advantages of applying this technique, we can verify that it allows us to solve several of the deficiencies presented by modern fuzzing. These deficiencies are presented below.

4.2.1. Fuzzing Different Components of the Connection. In general, the current fuzzing tools are only designed to verify some points of the connection. This means that if we use a tool to test a particular server, it cannot normally be used to repeat the process on a client, or at least not without investing a great deal of effort in modifying the structure of the framework. The solution to this problem marks one of the main characteristics of the tool that is being presented, since the objective to be achieved is to reduce the effort on the part of the user for verification of the security of applications that implement the MQTT protocol. With the proxy technique, because the fuzzer is in the middle of the communication, the main objective is the packages that circulate between the different components. Thus, the fuzzing tool is not built for a particular server or client, but for a given package set. Because the specification of the packages is standard for all applications that implement the protocol, the fuzzing process decouples completely from the point of the connection (client, broker, etc.) that is performing testing, focusing solely on the packages that are being exchanged.

4.2.2. Fuzzing Messages Based on Previous Responses. In some situations, it is not possible to apply fuzzing to certain packets in a protocol to determine whether the values are correctly processed by the target machine. This is because some of its fields are based on a previous message. If you want to test a particular value of a package that has a random handle field that has been previously sent by the server, it is not enough to establish a connection and continuously send this type of package, since they will be rejected for having an incorrect handle field, and the destination application would never get to process the value, and therefore the fuzzing process would not be carried out in any of the cases.

This is another problem solved with the proxy approach. Since messages that are filtered and processed by the fuzzer come from a legitimate client and broker that establish a legitimate connection, fields that have been previously sent from one end to the other will remain intact and with the proper value.

4.3. Template-Based Fuzzing. In this paper, we present a novel, template-based fuzzing technique that aims to solve the problems presented in Sections 2.1 and 2.2.

As explained earlier, the current fuzzing tools use an approach that tries to simplify work for users by recalculating the control fields automatically using the block-based technique. Even so, this method continues to prove highly complex for users who wish to perform security checks on a specific protocol. The code that is shown below represents all the sentences that are required to implement a small program that allows application of fuzzing to four messages in a very simple protocol (FTP) through a framework called *Boofuzz* [27], which is widely used nowadays, and the successor to *Sulley* [39], which in turn is heavily influenced by *SPIKE*. As can be seen, the definition of complex protocols in this type of framework is still a tedious task, in addition to requiring a thorough knowledge of the tool itself and the entire specification of the protocol. It is at this point that the template-based approach would be useful.

```
def main():

    session = Session(

        target=Target(
        connection=SocketConnection("127.0.0.1",
        8021,proto=\tcp")))

    s_initialize("user")
    s_string("USER")
    s_delim(" ")
    s_string("anonymous")
    s_static("\r\n")
    s_initialize("pass")
    s_string("PASS")
    s_delim(" ")
    s_string("james")
    s_static("\r\n")
    s_initialize("stor")
    s_string("STOR")
    s_delim(" ")
    s_string("AAAA")
    s_static("\r\n")
    s_initialize("retr")
    s_string("RETR")
    s_delim(" ")
    s_string("AAAA")
    s_static("\r\n")
    session.connect(s_get("user"))
    session.connect(s_get("user"),s_get("pass"))
    session.connect(s_get("pass"),s_get("stor"))
    session.connect(s_get("pass"),s_get("retr"))
    session.fuzz()
```

The template-based approach works as follows:

(i) The tool listens in the middle of the communication as if it were a sniffer, using the proxy technique. The user has previously had to provide a series of parameters whereby the packets that pass through it will be filtered. These are the fields that were discussed in previous sections.

(ii) The user generates traffic between the client and the legitimate server of the protocol that he or she wants to fuzz. When the packets that were specified in the previous point are intercepted by the tool they are filtered and processed.

(iii) After processing the package, a *.json* template is automatically generated with the following format.

This portion of the template shows the MQTT *Publish* layer of an MQTT package. As you can see, each of the fields in the package appears and two extra attributes are added to each: *fuzzable* and *recalculate*. All the user has to do to apply fuzzing to a particular field of a package is to modify the *fuzzable* attribute by assigning the value true. The user will also have to assign the value true to the recalculate attribute of the fields that are considered to be recalculated automatically in order to maintain packet consistency. The tool will automatically enter the verification values in the fields that have been marked as *fuzzables* and will also recalculate all the fields in the package that have the *recalculate* flag set to *true*.

As we can see in Figure 1, the generation time of the templates is reasonably fast, and the generation algorithm is $O(n)$, which means that the generation time remains constant, regardless of the number of templates generated.

Thus, the third problem is solved, since the user does not have to know any details of the structures used by the tool or the protocol itself, besides the fields to which he or she wants to apply fuzzing, and in any case, the fields that he or she wants to recalculate. Note that, in order to make modifications to the template, the user does not require any special tool; this can be edited with a common text editor, as long as the *.json* structure is maintained.

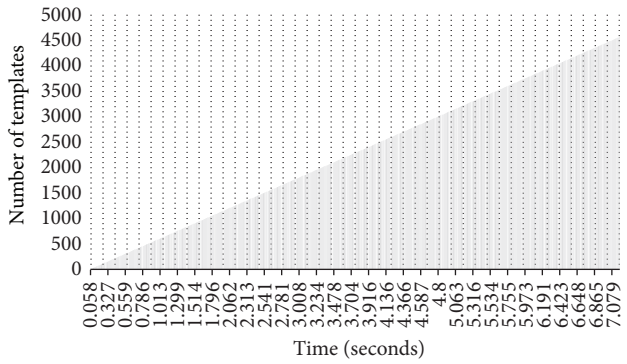


Figure 1: Time used by the application to generate templates.

4.4. Test Cases Generation. Generating the values with which an application is going to be tested is an important part of the fuzzing process [40, 41]. Often, the user is interested in running the test with a custom dictionary that has been generated with an external application; at other times, meanwhile, the user is interested in automatic generation of test cases with a certain degree of intelligence. The proposed tool implements both approaches and tries to maintain simplicity of use in both cases. The details of how this is implemented are discussed in the following section.

5. System Design

This section discusses the design considerations motivating our design and then describes the MQTT fuzzer system architecture.

5.1. Architecture of the Fuzzer. In this section, we will discuss the structure of the tool from the point of view of its design and implementation. We will take a tour of the main modules of which it is composed and its functionality. In addition, some secondary actions that the tool must carry out in order to make the fuzzing process satisfactory will be discussed.

Figure 2 shows the general architecture. The tool is composed of the following modules.

Mitmfuzzer. The *mitmfuzzer* module is the driver from which the rest of the application functions are called. Within it, the arguments that the user enters are parsed; this is done using the python module *argparse* [42]. In addition, it provides a small interface that shows the state of activity of the tool in a given situation.

Sniffer. The *sniffer* module is one of the main features of the tool. It is responsible for listening in the middle of a connection in order to filter and process the packages. It thus filters those specified by the user to subsequently generate a template. The core of the implementation of this module is based on *Scapy* [43], a framework for low-level treatment of network packets, which supports a large number of protocols. Once this module has detected a package selected by the user, it processes and serializes it in a certain format, which enables its processing using the python programming language. This package will be provided to the *template* module for template generation.

Template. This module receives a package in a certain format from the *sniffer* module and is in charge of processing it in order to generate the template in *.json* format. The generated templates are stored in a templates directory and will be used later by the fuzzer to identify the packages and fields to be fuzzed and recalculated.

Fuzzer. The fuzzer module is the most important module of the tool, as it performs the process of listening, packet filtering, generation, and insertion of test cases. The input of this module is a template file that must have been generated previously by the *template* module. Through the use of *iptables* and *nfqueue* [44, 45], the module continues listening to the communication as if it were a sniffer, redirecting the packages that are not identified with the template that has been introduced, and filtering and processing the packages that match the template. When a matching package is processed, all of its fields are compared to those in the template, looking to see whether one of them has been selected by the user to be fuzzed. In the case of one or more fields, the module checks whether the user has entered a custom dictionary to perform the test process: if so, the module will retrieve one of the test cases provided by the user and enter it in the field that was indicated for verification. Where the user has not provided a custom test case, the module will call *Radamsa* [46], passing as a parameter the file with the valid example case that is in *validcases/fieldnamedirectory*. Radamsa is a stock generator specially designed for software verification. It works by reading sample files that contain correct data, and through a series of algorithms it mutates this data, thus providing some intelligence, so that the generated results are more likely to lead to an error. *Radamsa* will automatically generate 50 different test cases and the module will take one of them to enter it in the field to be fuzzed. It should be noted that the generation of test cases performed by the module is continuous and infinite; when the module exhausts the 50 test cases generated, it automatically calls Radamsa to generate another 50 new cases.

Scapy. Scapy is a library for packet manipulation that supports a large number of network protocols. It has been considered necessary to name it in the application architecture, because it forms an important part of the core of the application. The advantage of Scapy, in addition to its extensive protocol support, is that it uses a block-based approach. This means that if you modify one of the fields in a package, you can recalculate the lengths and other control fields very simply and automatically. When a package arrives at the *Fuzzer* module, it sends it to Scapy for processing; Scapy then returns a structure that represents the package and which is easy to manipulate. After you have finished manipulating the MQTT package, Scapy takes this manipulated package, which is probably incorrect due to inconsistencies in control fields such as length fields (if text has been inserted or deleted) or checksum fields (if any byte of the package has been modified), recalculates all control fields using a block-based approach, and encapsulates the data as it was in the original

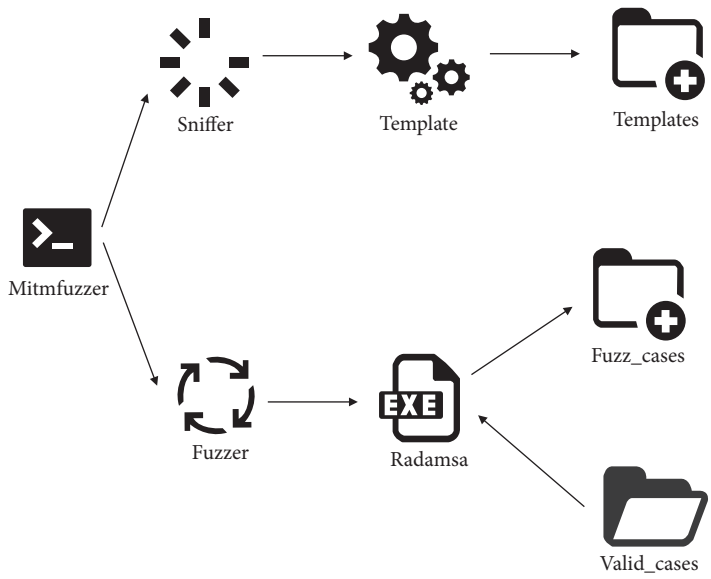


Figure 2: Architecture of the MQTT fuzzer.

package. It forwards this packet to the *Fuzzer* module, and the *Fuzzer* forwards it to the legitimate application. It is worth highlighting that Scapy did not have support for the MQTT protocol, and since it is the protocol object of study in this paper, we extended the library provided by adding support to MQTT. Currently, the module developed is part of the official repository of Scapy [47].

5.2. Test Cases Generation Implementation. In this section, we describe how the automatic generation of test cases has been applied in the implementation phase of the framework.

5.2.1. Automatic Generation of Test Cases. As explained in the description of the architecture of the tool, the automatic generation of test cases is carried out by an external application called Radamsa. This application is known to have been used to discover vulnerabilities like CVE-20073641 and CVE-2007-3644 (archive_read_support_format_tar.c library vulnerabilities), CVE-2008-6536 (7-zip program vulnerability), and CVE-2010-2482 (LibTIFF 3.9.4 vulnerability) among many others. The way in which the proposed tool uses this module is as follows: in the directory of the tool there are two important folders: a directory called *valid-cases*, composed of a set of subdirectories, one for each field of the package to be investigated. Inside these subdirectories, there are one or more sample files with correct data for that particular field. These will be provided to Radamsa to mutate them and generate the test cases. On the other hand, there is another directory called *fuzz-cases*, inside which a directory has been created for each of the fields to be fuzzed in a certain package. Radamsa automatically generates all field-specific test cases inside it, so that the tool subsequently retrieves them and inserts them into the packages.

5.2.2. Using Custom Test Cases. If, instead of using the automatic generation of test cases, it is desired to use a set of cases, generated either with another tool or manually, the user can do so in a straightforward manner by performing the following steps on the directory structure explained above:

(i) Inside the *fuzz-cases* directory, create a subdirectory with the exact name of the field you want to fuzz.

(ii) Inside the created subdirectory, enter all the test cases, one per file. The order or the name that is given to the file is not relevant.

6. Experimentation and Results

In this section, we present the results of applying the tool to a series of applications that are widely used today. All the test scenarios that are presented have been carried out in a controlled environment. The tools that are tested are open source and their use is free.

6.1. Performance Considerations. In this section, we have taken into account the performance implications of the tool. The section is divided into several subsections that evaluate the different functionalities of the presented tool and the impact of each of them on its performance.

6.1.1. Packet Processing. As has been presented in previous sections, the tool is located in the middle of the communication between a client and a broker, and from there it begins to modify all the network packets that flow between both, applying the proxy fuzzing technique. Because of this, much of the processing load of the tool corresponds to the modification and processing of packets on the fly, understanding processing such as insertion of test cases in the packets data fields and the recalculation of all the control fields of the previous layers.

Bearing this in mind, one of the aspects we have measured is the processing time per package. As can be seen in Figure 3,

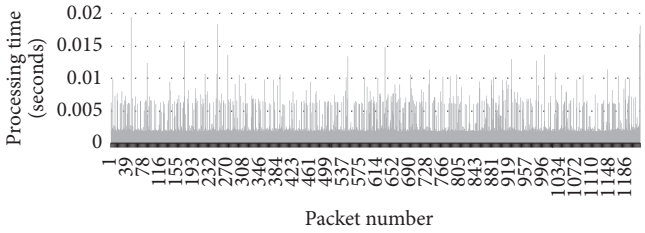


Figure 3: Processing time per package.

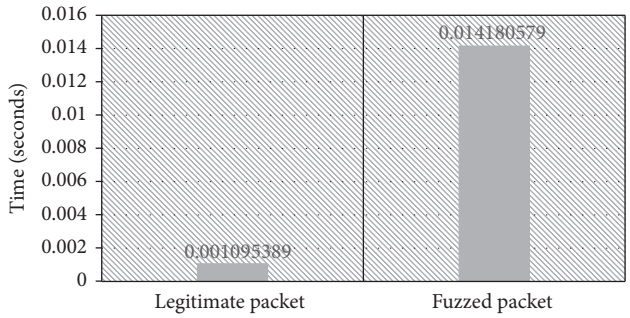


Figure 4: Difference between the transit time of a legitimate and a fuzzed package.

the processing time of each of the packages to which a test case is inserted remains relatively constant, with some variations due to the test case being inserted. If the test case has a longer length, the processing time will be longer because it will require recalculation of more fields. For the construction of the graph, a subset of 1300 network packets has been considered, which have reported an average processing time of 0.003699 seconds. This can be considered an acceptable time to keep the connection stable.

Once the processing time has been calculated for each packet, the arrival delay of a set of 100 packets after being processed has been calculated. This has been done because not only the overall processing time of a packet consists of inserting a test case and recalculating all the control fields of the lower layers, but also it is necessary to consider the delays caused by sending the packet from kernel space to user space so that it can be modified, sending of the package from user space to kernel space so that it can be sent, the additional time that it takes to be transported through the network, and so on.

As can be seen in Figure 4, the transit time of a fuzzed packet increases by approximately 90% with respect to the time of a legitimate packet, a total of 0.013085 seconds, which remains an acceptable time to maintain the connection stable without excessive delays.

6.1.2. Fuzzer Load and CPU Consumption. In previous sections, we have shown the number of templates that the tool is capable of generating and the time it takes to generate them. Another important performance measure is the number of test cases that the fuzzer is able to insert per time unit and the CPU consumption of the host machine.

The fuzzer has several customization features that allow you to select the time between test cases inserted, so that

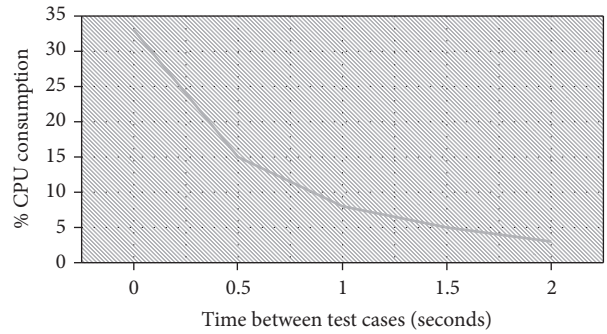


Figure 5: CPU consumption in relation to delay time between test cases.

you can insert everything as fast as possible or from time to time. In order to evaluate the CPU consumption of the host machine, different periods of time between inserted test cases have been taken into account.

As we can see in Figure 5, the CPU consumption of the machine that houses the fuzzer varies considerably depending on the delay time that is left between each inserted test case. This allows customizing the tool to be used in environments with fewer resources.

6.2. Application Scenarios. The term "application scenarios" refers to the possibilities offered by the tool within a connection to apply fuzzing to its elements. As explained in the previous sections, with the adopted approach it is possible to fuzz the different points of the connection of a protocol. The following are a series of use cases based on the MQTT protocol:

(i) Pub-fuzzer-broker-Sub: in this case, the tool would be placed between the client that is posting a message and the broker, in such a way that the tool could fuzz the messages that flow from the client to the server and the messages sent from the server to the client that is publishing.

(ii) Pub-broker-fuzzer-Sub: in this case, the scenario would change a little: the fuzzer would be between the broker and the client that is subscribed, waiting for the reception of messages. The tool could fuzz the messages from the broker to the client that is listening and the messages from the client to the broker, which will normally be acknowledgments.

6.3. Results. After applying the tool to some of the current brokers and clients, the fuzzer has been able to detect several failures that have led to denial of service and that may be potentially exploitable to perform other types of attack techniques. Some of these failures are as follows:

(i) Denial of service to the *MOQUETTE* broker v0.10 after the incorrect processing of a fuzzing package and throwing a Java exception that breaks the application

(ii) Error in handling the incoming connections by the broker *MOQUETTE* v0.10 after parsing a fuzzing package, which originates a connection reset

(iii) Denial of service of a *MOSQUITTO* client v1.4.11 that is subscribed to a certain topic when it receives a fuzzing message from the broker

This demonstrates that the fuzzing approach used provides real results in applications widely used by IoT devices around the world, and it can therefore be used as a security measure to ensure that devices in a given network meet minimum security standards.

7. Conclusion

The aim of this work was to contribute to improving the security of IoT devices and more specifically of the applications that implement a protocol widely used by Internet of Things (MQTT) as communication protocol for exchanging information. For this purpose, we developed a framework to perform security tests on MQTT. The tool implements a novel fuzzing technique based on templates, which according to our knowledge has not been used previously. The fuzzing technique presented here contributes to the field by improving some of the deficiencies of current fuzzers. The significant contribution of this framework is that it provides flexibility to fuzz the different points of a connection without making any adaptation effort. Among the other contributions explained during this paper, it is worth highlighting that it allows fuzzing packages that are based on a previously provided packet, and it facilitates portability and error reporting by exchanging templates. Finally, this technique simplifies the security analysis of the MQTT protocol to both users and applications by using the template-based approach, providing a way to fuzz the protocol without knowing or defining its specification.

Experimentation results gave acceptable processing time per fuzzed package. Moreover, it was observed that the tool behaves differently depending on the time that passes between each test case inserted. We were able to reduce the CPU consumption in the host machine to a minimum value of 2%. This flexibility to control the CPU consumption allows the use of this tool in environments with low processing power devices, such as Smart Cities. The tool was used to test vulnerabilities in widely used clients such as MOQUETTE or MOSQUITTO, with problems reported such as denial of service and communication resets of brokers. These discovered vulnerabilities probe the effectiveness of the tool.

The framework also has some limitations. The most significant one is related to the reporting and detection of errors. Error detection is carried out through the execution of the application to be tested under a debugger. Obviously, this needs to be automated to improve efficiency and usability. Another significant limitation is that the framework is currently only available for verifying MQTT protocol security; therefore the tool is not efficient for IoT architectures implementing several different protocols.

To improve the aforementioned limitations, as part of a future project, we will extend the framework to allow the verification of a wider range of network protocols used by IoT devices. Additionally, we are analysing the possibility of using the tool as a service that performs a security analysis of all the elements that are incorporated into a network for the first time. This would make it possible to ensure a minimum level of security and reliability for all the components of the infrastructure.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015.

[2] B. Chowdhury and M. U. Chowdhury, "RFID-based real-time smart waste management system," in *Proceedings of the In 2007 Australasian Telecommunication Networks and Applications Conference*, pp. 175–180, 2007, https://doi.org/10.1109/ATNAC.2007.IEEE.

[3] B. Padmavathi, "Implementation of IOT Based Health Care Solution Based on Cloud Computing," *International Journal of Engineering and Computer Science*, 2016.

[4] D. Gachet Páez, M. de Buenaga Rodríguez, E. Puertas Sánz, M. T. Villalba, and R. Muñoz Gil, "Healthy and wellbeing activities' promotion using a Big Data approach," *Health Informatics Journal*, p. 146045821666075, 2017.

[5] D. G. Páez, M. de Buenaga Rodríguez, E. P. Sánz, M. T. Villalba, and R. M. Gil, "Big Data Processing Using Wearable Devices for Wellbeing and Healthy Activities Promotion," in *Ambient Assisted Living. ICT-based Solutions in Real Life Situations*, vol. 9455 of *Lecture Notes in Computer Science*, pp. 196–205, Springer International Publishing, Cham, 2015.

[6] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, "Smart meters for power grid - Challenges, issues, advantages and status," in *Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition, PSCE 2011*, USA, March 2011.

[7] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.

[8] M. T. Villalba, M. de Buenaga, D. Gachet, and F. Aparicio, "Security analysis of an IoT architecture for healthcare," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 169, pp. 454–460, 2016.

[9] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1125–1142, 2017.

[10] M. B. Yassein, M. Q. Shatnawi, and D. Al-Zoubi, "Application layer protocols for the Internet of Things: A survey," in *Proceedings of the 2016 International Conference on Engineering and MIS, ICEMIS 2016*, mar, September 2016.

[11] H. Yang, Y. Zhang, Y.-P. Hu, and Q.-X. Liu, "IKE vulnerability discovery based on fuzzing," *Security and Communication Networks*, vol. 6, no. 7, pp. 889–901, 2013.

[12] M. Sutton, A. Greene, and P. Amini, *Fuzzing: brute force vulnerabilty discovery*, Addison-Wesley, Boston, Mass, USA, 2007.

[13] H. Yang, Y. Zhang, Y. Hu, and Q. Liu, "IKE vulnerability discovery based on fuzzing," *Security and Communication Networks*, vol. 6, no. 7, pp. 889–901, 2013.

[14] HiveMQ, "Enterprise MQTT Broker 2016," https://www.hivemq.com/wp-content/uploads/hivemq-product-sheet-v2-1.pdf.

[15] HiveMQ, https://www.hivemq.com/blog/mqtt-security-fundamentals-authenticationusername-password.

[16] I. Hedi, I. Špeh, and A. Šarabok, "IoT network protocols compa rison for the purpose of IoT constrained networks," in *Proceedings of the 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017*, pp. 501–505, Croatia, May 2017.

[17] T. Heer, O. Garcia-Morchon, R. Hummen, S. L. Keoh, S. S. Kumar, and K. Wehrle, "Security challenges in the IP-based Internet of Things," *Wireless Personal Communications*, vol. 61, no. 3, pp. 527–542, 2011.

[18] A. Esfahani, G. Mantas, R. Matischek et al., "A Lightweight Authentication Mechanism for M2M Communications in Industrial IoT Environment," *IEEE Internet of Things Journal*, pp. 1-1.

[19] R. Neisse, G. Steri, and G. Baldini, "Enforcement of security policy rules for the internet of things," in *Proceedings of the 2014 10th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2014*, pp. 165–172, Cyprus, October 2014.

[20] S. Shin, K. Kobara, C.-C. Chuang, and W. Huang, "A security framework for MQTT," in *Proceedings of the 2016 IEEE Conference on Communications and Network Security, CNS 2016*, pp. 432–436, USA, October 2016.

[21] A. Manzoor, "Securing Device Connectivity in the Industrial Internet of Things (IoT)," in *Connectivity Frameworks for Smart Devices*, Computer Communications and Networks, pp. 3–22, Springer International Publishing, Cham, 2016.

[22] J. Foster, V. Osipov, N. Bhalla, N. Heinen, and D. Aitel, "Buffer Overflow Attacks," *Buffer Overflow Attacks*, 2005.

[23] K. Kaspersky and A. Chang, "Remote code execution through Intel CPU bugs," in *Proceedings of the In Hack In The Box (HITB)*, Malaysia, 2008.

[24] D. Reynders, S. Mackay, and E. Wright, *Modbus overview. Practical Industrial Data Communications*, 10.1016/b978–3/50012-7, 2004.

[25] D. Serpanos and M. Wolf, "Security Testing IoT Systems," in *In Internet-of-Things (IoT) Systems*, pp. 77–89, Springer, Cham, Switzerland, 2017.

[26] D. Aitel, *The advantages of block-based protocol analysis for security testing*, Immunity Inc, February 2002.

[27] J. Pereyda, boofuzz: Network Protocol Fuzzing for Humans, http://boofuzz.readthedocs.io/en/latest/.

[28] G. Banks, M. Cova, V. Felmetsger, K. Almeroth, R. Kemmerer, and G. Vigna, "SNOOZE: Toward a Stateful NetwOrk prOtocol fuzZEr," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 4176, pp. 343–358, 2006.

[29] H. J. Abdelnur, R. State, and O. Festor, "KiF: A stateful SIP fuzzer," in *Proceedings of the 1st International Conference on Principles, Systems and Applications of IP Telecommunications, IPTComm '07*, pp. 47–56, USA, July 2007.

[30] S. Gorbunov and R. Rosenbloom, *AutoFuzz, Automated network protocol fuzzing framework*, Department of Mathematical and Computation Sciences, University of Toronto, Mississauga, Canada, 2010.

[31] D. Aitel, An Introduction to SPIKE, the Fuzzer Creation Kit, https://www.blackhat.com/presentations/bh-usa-02/bh-us-02-aitel-spike.ppt.

[32] "Github.org mqtt_fuzz," https://github.com/F-Secure/mqtt_fuzz.

[33] K. Mladenov, S. van Winsen, C. Mavrakis, and K. P. M. G. Cyber, Formal verification of the implementation of the MQTT protocol in IoT devices,.

[34] "OWASP.org, ZAP Proxy," http://www.zaproxy.org/.

[35] "PortsWigger.net, BurpSuite," https://portswigger.net/bur.

[36] "Github.com, ProxyFuzz," https://github.com/SECFORCE/proxyfuz.

[37] L. Landauer, "Fuzzing Requests And Responses Using A Proxy," U.S. Patent Application No. 11/276,454.

[38] OASIS.org, "MQTTVersion3.1.1:OASISStandard," http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html.

[39] A. Takanen, J. Demott, and M. Charles, "Fuzzing for Software Security Testing and Quality Assurance , Artech House Information Security and Privacy," *Fuzzing for Software Security Testing and Quality Assurance , Artech House Information Security and Privacy*, 2008.

[40] A. A. Sofokleous and A. S. Andreou, "Batch-optimistic test-cases generation using genetic algorithms," in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2007*, pp. 157–164, Greece, October 2007.

[41] R. Abbassi, S. Guemara, and F. El, "Towards a test cases generation method for security policies," in *Proceedings of the 16th International Conference on Telecommunications, ICT 2009*, pp. 41–46, Morocco, May 2009.

[42] "Python.org, Argparse," https://docs.python.org/3.4/library/argparse.html.

[43] "Scapy A Python Tool For Security Testing," *Journal of Computer Science & Systems Biology*, vol. 8, no. 3, 2015.

[44] J. Alan, "Netfilter and IPTables - A Structural Examination," *SANS Institute*, 2004.

[45] "Netfilter.org, Netfilter," https://www.netfilter.org/.

[46] "University of Oulu, Radamsa," https://www.ee.oulu.fi/roles/ouspg/Radams.

[47] "Scapy.org, MQTT layer for Scapy," https://goo.gl/oo45XC.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 1073216, 13 pages
https://doi.org/10.1155/2018/1073216

Research Article

TrustyFeer: A Subjective Logic Trust Model for Smart City Peer-to-Peer Federated Clouds

Heba Kurdi,[1,2] **Bushra Alshayban**,[1,3] **Lina Altoaimy**,[4] **and Shada Alsalamah**[5,6]

[1]*Computer Science Department, King Saud University, Riyadh, Saudi Arabia*
[2]*Mechanical Engineering Department, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA*
[3]*Technical and Vocational Training Corporation, Riyadh, Saudi Arabia*
[4]*Information Technology Department, King Saud University, Riyadh, Saudi Arabia*
[5]*Information Systems Department, King Saud University, Riyadh, Saudi Arabia*
[6]*Media Lab, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA*

Correspondence should be addressed to Heba Kurdi; hkurdi@ksu.edu.sa

Received 16 November 2017; Revised 15 January 2018; Accepted 24 January 2018; Published 25 February 2018

Academic Editor: Syed H. Ahmed

Cloud computing plays a major role in smart cities development by facilitating the delivery of various services in an efficient and effective manner. In a Peer-to-Peer (P2P) federated clouds ecosystem, multiple Cloud Service Providers (CSPs) collaborate and share services among them when experiencing a shortage in certain resources. Hence, incoming service requests to this specific resource can be delegated to other members. Nevertheless, the lack of preexisting trust relationship among CSPs in this distributed environment can affect the quality of service (QoS). Therefore, a trust management system is required to assist trustworthy peers in seeking reliable communication partners. We address this challenge by proposing TrustyFeer, a trust management system that allows peers to evaluate the trustworthiness of other peers based on subjective logic opinions, formulated using peers' reputations and Service Level Agreements (SLAs). To demonstrate the utility of TrustyFeer, we evaluate the performance of our method against two long-standing trust management systems. The simulation results show that TrustyFeer is more robust in decreasing the percentage of services that do not conform to SLAs and increasing the success rate of exchanged services by good CSPs conforming to SLAs. This should provide a trustworthy federated clouds ecosystem for a better, more sustainable future.

1. Introduction

Smart cities hold the promise for economic development, organizational performance, social equity, and quality of living. This attracts the majority of the world's population to seek a sustainable future by moving to urban environments [1]. Consequently, contemporary cities are left with no option but to utilize their infrastructure and services to handle such increasing demand. City authorities need innovative solutions than can efficiently address the problems of urban development. The vision of a smart city is to deploy Information and Communications Technology (ICT) in a smart and efficient manner to enhance the quality of life for citizens and ensure access to all the services offered by the city's government [1, 2].

Cloud federation is an emerging approach in the modern cloud computing landscape that can be adopted to achieve the vision of the smart city [1]. Advances in cloud computing have made it the first choice for storing data and providing functionality, including, but not limited to, pay-per-use, on-demand services, elasticity, remote access, and the pooling of resources [3]. To ensure a high Quality of Services (QoS), services offered over the cloud are based on various Service Level Agreements (SLAs) between a Cloud Service Provider (CSP) and cloud service consumer [4]. However, it is very important for CSPs to guarantee and maintain the quality of their services at satisfactory levels, and thus, when a CSP is asked for any unavailable service, it may interact with other providers to offer this service to conform to the SLA. This is precisely the notion behind federated

clouds. The term federated cloud refers to a type of cloud computing paradigm where several independent CSPs agree to share their infrastructures/resources to accommodate a wider range of requested services with a higher QoS [5]. Each participant in the federated cloud model has access to a much larger pool of resources, and the peak-load handling capacity for each cloud increases without having to administer or maintain additional computing resources. There are two categories of federated clouds based on the communication/interaction methods: Centralized and Decentralized Peer-to-Peer (P2P) [6, 7]. On the one hand, in the centralized category, the system relies entirely on a central party that acts as a control/connection node between different clouds [8]. In the decentralized P2P category, on the other hand, each cloud is linked directly to the other clouds within the cloud ecosystem without central management [9]. Nevertheless, due to the open nature of federated cloud systems, distributed structure, and independent provisioning of resources, CSPs in these systems share resources with each other without a preexisting trust relationship. Therefore, they are vulnerable to harmful attacks by selfish and malicious CSPs.

Establishing a trust relationship between CSPs is one of the key challenges in a P2P federated cloud environment, thus leading to it becoming an emerging area of research [10]. Malicious CSPs can harm P2P cloud federation in several ways that can hinder the overall effectiveness of such federations. First and foremost, malicious CSPs can provide inauthentic services or services that do not conform to SLA. Second, they may also lie about their feedback regarding the QoS they receive from other service providers [11]. Finally, such malicious CSPs with low reputation values can create new accounts to clear their already damaged reputations and continue harming the system; this is known as whitewashing [12]. Hence, establishing a trust relationship between CSPs is one of the key challenges in a P2P federated cloud environment. According to Filali and Yagoubi [13], trust measures and metrics are two main parts that must be considered when establishing a trust management system. On the one hand, trust measures have values that are either discrete or continuous. Metrics (mathematical model), on the other hand, can be based on various approaches: simple summation, average of ratings [14], flow-based models [15, 16], fuzzy logic [17–21], probabilistic models such as Bayesian systems [22–24] or beta probability density [25, 26], or a subjective logic approach [27, 28].

Given the need to establish trust between CSPs in P2P federated clouds, we propose TrustyFeer, a trust management system that implements a trust overlay network (TON) [29] on top of the federated cloud and focuses only on the interactions between P2P CSPs. We also compare the results against two state-of-the-art reputation systems, EigenTrust [30] and TNA-SL [28]. The reason for choosing these two systems is because they are the most known trust management systems and many recent proposed systems were based on them [31–35]; in addition, they are already implemented in the simulation tool [36] that we used in our evaluation. Based on the results, we found that our proposed system achieves a better performance in varying the number of CSPs and malicious CSPs in the system. The main contributions of this paper can be summarized as follows:

(1) We applied the basic idea of subjective logic [27, 28] and propose new trust formulas (as described in Section 3.3) and weight trust values based on the global reputation of each CSP and the degree of its conformance to an SLA.

(2) We built a strictly controlled evaluation framework for simulating trust relationships in P2P federated clouds.

The remainder of this paper is organized as follows. Section 2 reviews the existing work on trust management systems for both P2P networks and cloud computing. Section 3 explains the design of the proposed system, while Sections 4 and 5 describe the evaluation plan and the experimental results. A summary and concluding remarks are provided in Section 6.

2. Literature Review

This section discusses several studies in the field of trust management systems in open environments, such as e-commerce, P2P networks, web services, grid computing, and cloud computing [50, 51]. For the purpose of this review, trust management systems are classified into two categories: trust systems in P2P networks and trust systems in P2P clouds.

2.1. Trust in P2P Networks. Trust management systems in P2P environments can be divided into three main categories. The first category is peer-based trust systems, where the decision for which source to download is reached based on the honesty of peers; secondly, we have file-based trust systems, where the decision for which source to download is reached based on the quality of the provided file; the final category is hybrid-based systems, which are combinations of peer and file systems [49]. Among the different categories, peer-based systems are the most commonly employed.

EigenTrust [30] is one of the most popular peer-based algorithms in P2P networks. It lessens the impact and influence of malicious nodes within the P2P system. In EigenTrust, the global reputation of a peer in the system is calculated using the left principal eigenvector of a matrix of normalized local reputation values. In addition, an overall history of the system is available, and each peer is known and considered in the calculation of the reputation values, which is performed in a distributed and node-symmetric manner with minimal overhead on the network [31].

An enhanced EigenTrust algorithm, called HonestPeer, was proposed in [35]. In this technique, the set of honest peers with high reputation values are given greater roles in calculating the global reputations of other peers. HonestPeer dynamically selects the honest peers based on the quality of the provided files, rather than depending only on the static group of pretrusted peers, as in EigenTrust. Compared with the EigenTrust algorithm, HonestPeer has a better success rate and a minimal percentage of inauthentic downloads.

In PeerTrust [37], certain factors are considered in the trust assessment of a specific node, such as the feedback

and its scope, the credibility of the source, the context of transactions, and the community context. The metric of general trust combines the above factors and then significantly decreases ordinary threats, such as man-in-the-middle attacks, nodes in compromised bases, and tainted information being spread within the decentralized environment of the P2P. However, in this scheme, the underlying presumption is that the trust value of a peer is a measure of its reliability. Therefore, peers with higher trust values always provide more reliable feedback, but this is not always correct.

The Grid Peer [52] trust model is another upgraded version of the PeerTrust model. This model has corrected certain flaws in the PeerTrust model and has also created a place for PeerTrust within the environment of the grid. This is achieved by modifying the definition of satisfaction criteria and introducing a decay function in the algorithm. The trust factor in the satisfaction criteria is responsible for handling all the requirements that can be satisfied by the resource source. The provider of resources is chosen from a performing grid by taking into account the basic requirements that the resource consumer wants to fulfill. The addition of a decay function, which is upgraded via a feedback trust calculation algorithm, forms the basic concept of the algorithm.

In CuboidTrust [39], three trust parameters, namely, contribution, trustworthiness, and quality of resource, are used to build four relations, and the global trust value of every single peer is calculated using the power iteration. CuboidTrust provided good results and brought about a substantial reduction in the number of inauthentic resource downloads in different threat models. The parameter of trustworthiness was considered in CuboidTrust and PeerTrust, and not in EigenTrust. Therefore, both CuboidTrust and PeerTrust perform efficiently, even in presence of various malicious peers in the system.

PowerTrust [38] utilizes a TON to simulate the interrelationships of trust present between nodes. This scheme utilizes the power-law feedback properties of P2P networks and uses random selection power nodes, which are those with the best reputations. Compared with the previously mentioned systems, PowerTrust achieved much better results in terms of the precision of global reputation and the speed of aggregation. However, PowerTrust is prone to threats by malicious pretrusted peers. This is because, in such a model, the status of completely reliable peers (pretrusted peers are considered to be in EigenTrust) is given to power nodes [53]. Therefore, there is a high probability of severe system damage if power peers turn malicious.

GossipTrust [41] is especially designed for unarranged P2P networks. The system leverages a gossip-based protocol to aggregate the score of a global reputation. Every peer randomly contacts the others and exchanges information regarding the reputation of the data in a periodic manner. The proposed gossip-based protocol is simple, does not require error recoveries, and provides controlled overheads compared with optimum deterministic protocols, which include information building of a data dissemination tree. In addition, GossipTrust provides a fast aggregation module of local trust scores, a new efficient scheme for storing reputation information, and secure communication using identity-based cryptography.

In [40], the proposed BP/P2P system computes reputation and trustworthiness values by using a belief propagation-based distributed message, passing algorithm between peers on a factor graph representation of a P2P network. Using BP/P2P, the reputations of peers are determined based on the quality of service a peer receives, and trustworthiness is determined based on the ratings provided by each peer after successful transactions. A comprehensive evaluation showed that the BP/P2P is efficient in calculating trust values, filtering malicious ratings, and reducing errors in the reputation values of peers. Moreover, compared with EigenTrust and PowerTrust, BP/P2P is more efficient in detecting and eliminating malicious nodes.

In VectorTrust [43], a Bellman–Ford-based algorithm is utilized for the quick compilation of trust scores. To analyze and compile trust values, the trust vector aggregation (TVAA) is proposed. Every single trust path in TVAA is collected, and the highest trust rating is given to a target peer. VectorTrust can be employed in decentralized and distributed networks, where no global trust data is available. As the complexity of both topology and P2P networks is increasing, VectorTrust scales effectively owing to its high speed of convergence and manageable costs.

Hierarchical fuzzy trust management (HF Trust) [42] makes use of fuzzy logic to model trusts. Every peer keeps records of all local exchanges, to determine if the peer has fulfilled the requirements or not. The factors related to trust are subjected to fuzzy inference by peers to produce a local trust index. All the data from every peer regarding local transactions is compiled by the HF Trust system, and every peer's global reputation is prepared. The seven significant parameters for evaluating trust are explained using an application that allows file sharing between peers. A considerable improvement in the performance of the P2P system has been demonstrated using this trust model, as it brings about a substantial reduction in the number of nongenuine files in the network.

The most important algorithm, on which the TrustyFeer system is based, is the trust network analysis with subjective logic (TNA-SL) algorithm [27, 28]. This algorithm is used to discover trust networks between two parties and derive trust measures from such networks. Trust in TNA-SL is stored as an opinion, and each opinion consists of four tuples (b, d, u, α). These four tuples represent belief, disbelief, uncertainty, and a base rate, respectively, where $(b+d+u) = 1$, and α (a real number) in the range $[0 \cdots 1]$. Subjective logic offers different types of operators from binary logic and probability calculus as well as specific operators for combining and merging different opinions. This variety of operators makes it possible to support a wide range of different applications and systems. Details regarding the opinion formulation and operators are addressed in the following sections.

The previously mentioned systems are categorized as peer-based approaches. There is less research available regarding file-based approaches. The system proposed in [48] is one such system, which depends on the quality of files when calculating trust values. This system prevents inauthentic

message distributions and downloads in a structured P2P network. It depends on a distributed hash table in which there is an ID key for each file-based on its name and contents. The reputation values are stored in the file repository and the peer repository of each peer. Only trustworthy peers can evaluate the files and give reputation values. The trustworthiness of a file is based on some system parameters and equations, which are calculated after each transaction.

AuthenticPeer [49] is a hybrid-based trust management system in P2P wireless sensor networks. It uses techniques from the peer and file approaches to maximize features and overcome limitations. An evaluation of this system showed that the performance of the reputation system is enhanced compared with EigenTrust and Incremental EigenTrust.

2.2. Trust in P2P Cloud Computing. The volume of research on the P2P cloud computing environment is quite limited. The most similar system to the TrustyFeer approach is proposed in [46]. This system uses the feedback aggregation from distributed peer clouds in the system to calculate trust values. Peers are rated based on their previous compliance to SLAs with other peers. Penalties are employed to identify previous negative interactions, while rewards are used to identify previous positive interactions. However, this system only considers limited models of malicious peers and is evaluated under PeerSim, which is a P2P networks simulator.

In this model [44], there are two trust tables for each peer: a direct trust table and a recommended list. When a peer needs to calculate the trust value of another peer, it first checks its direct trust table. If a trust value exists for the peer, then it will adopt this. Otherwise, the peer checks the recommended list to find a peer that has a direct trust relationship with that specific peer. The trust value is determined based on the queries exchanged among nodes in the cloud. It also considers the following metrics: processing capacity, operating system, storage space, and links. However, this system is designed for private clouds and does not consider SLAs.

A reputation management model implemented for P2P in a cloud service provisioning environment was introduced in [45]. This model is similar to the one in [44] but makes use of Cloud Brokers (CBs). In this scheme, the CB is responsible for validating the trust based on certain parameters: storage capacity, processing capacity, links, and data cost. When one peer needs to communicate with another, it requests information from the CB or other peers. The system was validated using the CloudAnalyst tool [54]. However, in this system, the SLA measurements are also not considered.

Another trust management system in cloud computing has been designed especially for federated clouds [47]. This is a distributed framework that allows providers to determine the trustworthiness of different federated cloud computing providers. Trust is specified using personal experiences, reputation, and honesty ratings. The storage of trust values is distributed in each cloud, to allow clouds to make independent decisions regarding selection based on the trustworthiness of other clouds. Malicious peers in the system will have less ability to interfere with network operations. Therefore, providers can defend themselves against malicious ratings and satisfy the clients' QoS requirements. However, SLA measurements are not considered and information about the performance and simulation is not provided.

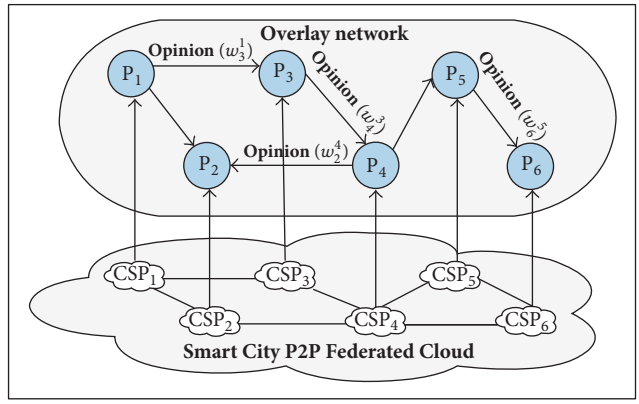


Figure 1: Trust overlay network over a P2P federated cloud.

In contrast to the above proposed schemes, our proposed system combines the advantages of the various trust management systems. First, it uses the global reputation weight when calculating trusts, as applied in PowerTrust [38]. Second, it builds a TON on top of P2P federated clouds, as in PowerTrust [38] and VectorTrust [43]. Third, it calculates trust and reputation values based on subjective logic, similar to [28], which is more accurate owing to the different factors it considers for each score. Fourth, it introduces new formulas for calculating the subjective logic components (i.e., belief, disbelief, uncertainty, and the base rate) based on the SLA parameters. Fifth, it uses the SLA parameters to calculate the trust, which are the most important parameters for improving the performance in cloud computing environments. Table 1 summarizes the different characteristics of the previously discussed systems.

3. System Design

TrustyFeer reputation system builds a TON [29] on the top of a P2P federated cloud, as shown in Figure 1. The vertices in the network ($CSP_1, CSP_2, \ldots, CSP_6$) represent the CSPs in the system, while the directed edges represent the direct recent transactions or interactions between two providers. An edge label represents the opinion of the source CSP regarding the destination CSP.

In this section, we describe the architecture and operations of the TrustyFeer system. We also discuss the computation of the local and global trust values using subjective logic [28].

3.1. System Architecture. The architecture of the TrustyFeer trust management system, illustrated in Figure 2, consists of three basic components, and each component has a different task. The basic components of the TrustyFeer trust management system are the system registry, the reputation database management system, and the provider peer. Next, we describe each component in detail.

(1) System registry: this is an in-memory database that contains the list of CSPs in the system and the services

Table 1: Comparison of different trust management systems.

Type	Systems	Environment	Tool used	Trust value evaluation	Performance measured
Peer-based trust systems	EigenTrust [30]	P2P network	P2P simulator	Calculating the sum of positive and negative ratings	Number of inauthentic files
	PeerTrust [37]	Decentralized P2P network	Mathematica 4.0	Normalizing the transaction rating	Trust computation error against malicious behaviors
	PowerTrust [38]	P2P Grid computing	Discrete-event driven simulator	Bayesian theory	Reputation convergence overhead, ranking discrepancy, and aggregation errors
	HonestPeer [35]	P2P network	P2P trust simulator	Calculating the global reputation of other peers	Success rate and the percentage of inauthentic downloads
	CuboidTrust [39]	P2P network	-	Using power iteration	Number of inauthentic resource downloads under various threat models
	BP-P2P [40]	P2P network	MATLAB	Using belief propagation distributed message	Error in the reputation values of peers, the computational complexity, and the communication overhead
	GossipTrust [41]	P2P network	P2P trust simulator	Using gossip protocol	Query success rate
	HFTrust [42]	P2P network	Simulation using Java	Using fuzzy logic	Number of inauthentic files in the system
	VectorTrust [43]	P2P network	VTSim simulator using Netlogo	Using Bellman-Ford-based algorithm	Convergence speed, communication overhead, malicious peer detection rate and detection speed
	TNA-SL [28]	P2P network	-	Using subjective logic	-
	Trust model for reliable file exchange [44]	P2P private cloud	-	Calculating the sum of all metric values multiplied by a given weight	-
	Reputation-based trust model [45]	P2P private cloud	CloudAnalyst environment tool.	Using a mathematical model	Trust rate
	Trust modeling [46]	P2P clouds	PeerSim simulator	Using game theory	Level of trust in terms of trusted and untrusted peers
	Service trust worthiness [47]	Intercloud computing	-	Based on quorum	-
File-based trust systems	Reputation management system [48]	Structured P2P network	-	Using file reputation and peer reputation	Degree of preventing untrustworthy files from spreading compared to existing systems
Hybrid-based trust systems	AuthenticPeer [49]	P2P sensor network	P2P trust simulator	Using [30] for peer quality and [48] for file quality	Success rate and the failure rate of authentic files

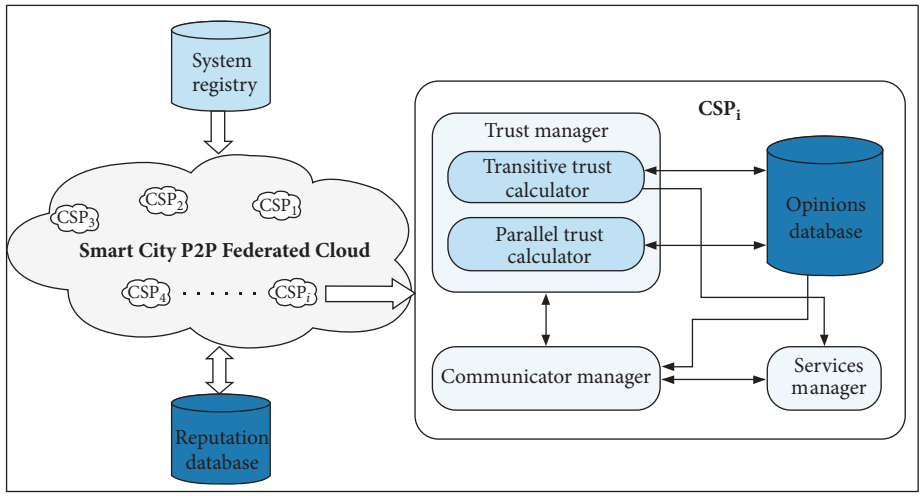


Figure 2: System architecture.

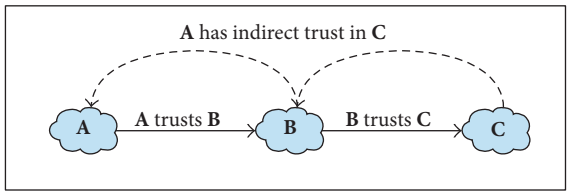


Figure 3: Transitive trust example.

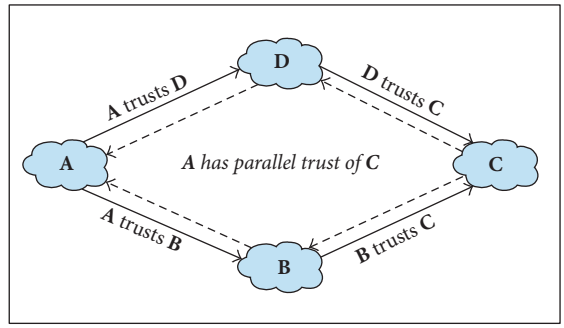


Figure 4: Parallel trust example.

they offer. It acts as a resource management system, which maintains traces of all providers in the system and dynamically updates the current services offered by each CSP.

(2) Reputation database management system: this stores the reputation matrix and updates the matrix after each round.

(3) Provider peer: each provider peer consists of the following four components:

(a) Trust manager: this is responsible for trust calculations. It receives the services score from the service manager to compute the trust matrix and sends this to the opinions database. It contains two components: the *transitive calculator* and the *parallel calculator*.

(i) The *transitive trust calculator* is used to calculate the indirect trust, as illustrated in Figure 3.

(ii) The *parallel trust calculator* is used to calculate the parallel trust for one node, as illustrated in Figure 4.

(b) Communication manager: this requests the system registry to check the services of other providers and update their available services. It also requests for other providers in the system to receive services. In addition, it updates the trust matrix for the client provider. It also checks the reputation database to obtain the trust values of other providers. In addition, it exports the trust matrix to the reputation database.

(c) Opinions database: this contains the matrices of opinions for all providers in the system. It also receives requests from the trust manager to store or update the matrix after each cycle.

(d) Services manager: this contains the list of services and accepts requests for services. In addition, it rates and normalizes the received services. It also sends the scores of rated services to the trust manager.

3.2. System Operations. The TrustyFeer system consists of a number of CSPs in P2P formation. Each provider peer in the system has an associated $n \times n$ matrix of opinions, where n is the number of providers in the system. In a similar approach to [28], the trust of provider A in provider B is represented by opinions on four factors, b, d, u, and α, which represent belief, disbelief, uncertainty, and the base rate, respectively, with $(b + d + u) = 1$. Here, α is a real number in the range $[0 \cdots 1]$.

Table 2 details the calculations of opinions, as proposed in [28], based on previous interactions. The new proposed

Table 2: Opinion calculation [28].

Belief	*(Positive/(positive + negative + 2.0))*
Disbelief	*(Negative/(positive + negative + 2.0))*
Uncertainty	*(2.0/(positive + negative + 2.0))*
Base rate	*1.0 if the CSP pretrusted, 0.5 otherwise.*

```
(1) procedure Initialization
(2)     for each provider in the system do
(3)         believe = 0.0
(4)         disbelieve = 0.0
(5)         uncertainty = 0.0
(6)         alpha = 0.0
(7)     end for
(8) end procedure
```

Algorithm 1: Initialize global trust matrix.

formulas for calculating belief, disbelief, uncertainty, and base rate values are discussed in the next section.

The discount operator is used to evaluate transitive chains. For example, when CSP A wants to calculate an opinion regarding CSP C using information from the intermediate CSP B, the discount operator is used. In this case, the notation is written as

$$\omega_C^{(A:B)} = \omega_B^A \otimes \omega_C^B. \tag{1}$$

The consensus operator is used to average two opinions together. For example, when CSP A and CSP B both have opinions about CSP C, the consensus operator is used to consolidate them. The notation for this is

$$\omega_C^{A\Diamond B} = \omega_C^A \oplus \omega_C^B. \tag{2}$$

Table 3 presents the calculation of the above two operators. In addition, the single trust value extracted from an opinion factor is called the expected value (E) and is calculated as in (3):

$$E = b + \alpha * u. \tag{3}$$

3.3. Trust Evaluation. At the initial stage, all CSPs in the system are initialized at the following values of opinion factors: $b = 0$, $d = 0$, $u = 1$, and $\alpha = 0.5$, as presented in Algorithm 1.

Then, after each transaction, the values b, d, u, and α of the opinion factors for the sender provider are updated using new proposed formulas (4), which consider the global reputation of the sender provider when calculating the opinion factors:

$$b_B^A = S_B^A * E$$

$$d_B^A = \left(1 - S_B^A\right) * E$$

Table 3: Calculation of discount and consensus operators [28].

Discount: $\otimes$	Consensus: $\oplus$
$b_c^{A;B} = b_B^A b_C^B$	$b_c^{A\Diamond B} = b_C^A u_C^B + b_C^B u_C^A$/denominator
$d_C^{A;B} = b_B^A d_C^B$	$d_C^{A\Diamond B} = d_C^A u_C^B + d_C^B u_C^A$/denominator
$u_C^{A;B} = d_B^A + u_B^A + b_B^A u_C^B$	$u_C^{A\Diamond B} = u_C^A u_C^B$/denominator
$\alpha_C^{A;B} = \alpha_C^B$	$\alpha_C^{A\Diamond B} = \alpha_C^A$

Denominator = $u_C^A + u_C^B - u_C^A u_C^B$.

$$u_B^A = 1 - \left(b_B^A + d_B^A\right)$$

$$\alpha_B^A = E^{t-1},$$

where t represents the current transaction. (4)

S_B^A is the score of the service given by provider A (i.e., the sender provider) to provider B (i.e., the receiver provider). Algorithm 2 shows the pseudocode for updating opinion values.

3.4. Trust Aggregation. Considering the scenario in Figure 5, where the five CSPs, A, B, C, D, and E, have opinions between them shown on edges, the procedure for determining the trust A has in E, described in Algorithm 3, is as follows:

(1) Create a matrix, OF, of opinion factors with dimension (5×5), where each element of the matrix defines a CSP–CSP relationship. All opinions are computed based on the formulas described in Section 3.3.

(2) Compute the square of the matrix OF using the discount and consensus operators, as presented in Table 3. The discount operator is used to calculate the trust on each path, and then the consensus operator is used to determine the consensus of these all together.

(3) Create a different matrix OF$'$ to store the opinions with maximum confidence resulting at each position throughout the multiplication process.

(4) Define a global trust matrix EV(OF$'$) that contains the expected trust values, where EV(OF$'A, E$) represents the trust the CSP A has in the CSP E.

4. Evaluation Methodology

The main factors that affect P2P federated clouds are the number of CSPs and the models and percentages of malicious CSPs. In order to evaluate the efficiency of our proposed system, we used the following materials:

(i) Hardware: processor: Intel core i5; speed: 1.1 GHz; RAM: 2 GB; hard disk: 200 GB; keyboard: standard keyboard; monitor: SVGA.

(ii) Software: to develop and test trust and reputation algorithms in dynamic environments, we used P2P trust simulator [36], which is a Java-based, well-established open source simulation framework. It consists of two main elements: the trace generator and

```
(1) procedure Update
(2)   if the client provider in a new direct transaction then
(3)     Update opinion values of server provider in the client provider's matrix
(4)     believe = ServiceScore ∗ ExpectedValueOfClientProvider
(5)     disbelieve = (1.0 − ServiceScore) ∗ ExpectedValueOfClientProvider
(6)     uncertainty = 1.0 − (believe + disbelieve)
(7)   end if
(8) end procedure
```

Algorithm 2: Update opinion values for server provider.

```
(1) procedure ComputeTrust
(2)   create matrix OF of opinion elements
(3)   for each CSPs in the system do
(4)     discount and then consensus opinions in OF
(5)   end for
(6)   for each position in matrix OF do
(7)     if the believe of the opinion in OF is max then
(8)       store the opinion with the max believe in OF′
(9)     end if
(10)  end for
(11)  export the stored value in OF′ to the global trust matrix
(12)  compute opinion expected values as trust values
(13) end procedure
```

Algorithm 3: Compute trust value.

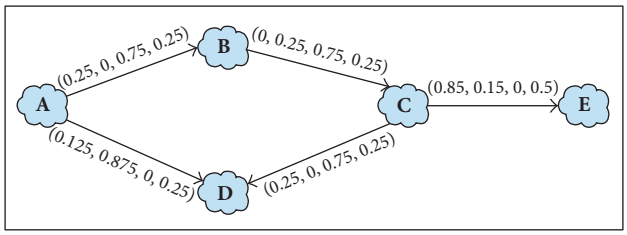


Figure 5: Example of a trust network graph of five CSPs.

the simulator. The trace generator outputs a trace file containing the initialization of the network and transactions. The simulator takes the output trace file and the reputation algorithm as input and then simulates the network run and outputs statistics on how the trust manager performed. The two benchmark systems, EigenTrust [30] and TNA-SL [28], are already implemented in the simulator and used in the evaluation of the proposed system. In addition, P2P trust simulator is used to evaluate trust and reputation systems in P2P network and clouds environments [35, 41, 46].

We considered the following performance measures:

(1) The percentage of inauthentic services exchanged by good CSPs, which should be minimized. These are the services that do not conform to SLA measures.

(2) The success rate for good CSPs, which describes the number of valid services received by good CSPs satisfying SLA measures divided by the number of transactions attempted by good CSPs.

$$\text{Success rate} = \frac{\#\text{ of valid services received by good providers}}{\#\text{ of transactions attempted by good providers}}. \quad (5)$$

The hypothesis is that using TrustyFeer in a P2P federated cloud will increase the success rate of services exchanged by good CSPs that conform to SLAs and decrease the percentage of invalid services that do not conform to SLAs under various conditions, such as different of models and percentages of malicious providers and different numbers of CSPs.

To evaluate the performance of the TrustyFeer trust management system in a P2P federated cloud environment, we simulated a representative sample of federated clouds of different scales and conducted a well-controlled experiment using the parameters shown in Table 4.

We identified the critical elements inherent in the design of a trust management system, which are the number of CSPs and the percentages and models of malicious CSPs and controlled their values as follows:

(1) Number of CSPs was selected in the range $[20 \cdots 70]$.

(2) Percentage of malicious CSPs was selected in the range $[10 \cdots 60]$.

Table 4: The simulation parameters.

Simulation parameter	Values
Application model	File sharing
Network model	Closed network
Numbers of CSPs	20, 30, 40, 50, 60, 70
Percentages of malicious CSPs	10%, 20%, 30%, 40%, 50%, 60%
Number of transactions	750

(3) Models of CSPs were four different models:

(a) Good providers, which provide honest feedback and a rate of authentic services between 90% and 100%, as we cannot expect good providers to be ideal ones.

(b) Purely malicious providers that provide inauthentic services or services that do not conform with SLAs or that lie about the feedback on the services they have received from other CSPs regarding the service quality (whether or not they conform to SLAs).

(c) Malicious feedback by providers that conform to SLAs but consistently lie about the quality of the services they receive.

(d) A malicious collective consisting of a group of cooperating malicious CSPs.

The extraneous variables, such as the number of transactions, were controlled using randomization to ensure a representative sample in all experiments. Two outstanding reputation algorithms, EigenTrust [30] and TNA-SL [28], were selected as benchmarks. The following two sets of experiments were designed, as shown in Table 5:

(1) In the first set, the percentage of malicious CSPs was constant at 20%, while the values for the number of CSPs were selected in the range of 20, 30, 40, 50, 60, and 70 CSPs.

(2) In the second set, the number of CSPs was constant at 40, while the percentage of malicious CSPs was selected in the range of 10%, 20%, 30%, 40%, 50%, and 60% malicious CSPs.

For each set, two mixed models of malicious CSPs were used, with two experiments taking different strategies: *naïve* and *collective*. Moreover, to increase the accuracy of this experimental study, we repeated each experiment ten times and calculated the average of the outcomes. The number of transactions was constant at 750.

5. Results and Discussion

A summary of the results of running 750 transactions for all sets of experiments is shown in Figures 6–13. Each set of experimental results represents an average of 10 simulation runs. As discussed previously, we used the success rate and percentage of invalid services to evaluate the efficiency of our proposed system.

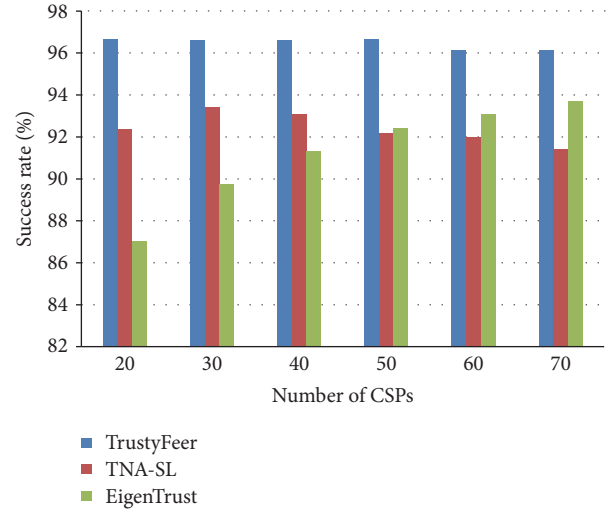


Figure 6: Success rate when different numbers of CSPs are considered (collective).

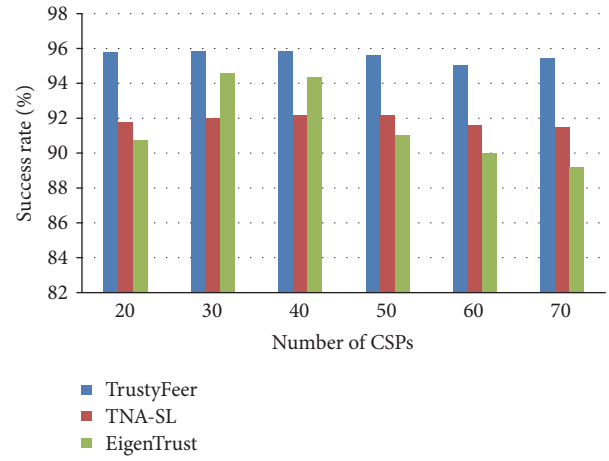


Figure 7: Success rate when different numbers of CSPs are considered (naïve).

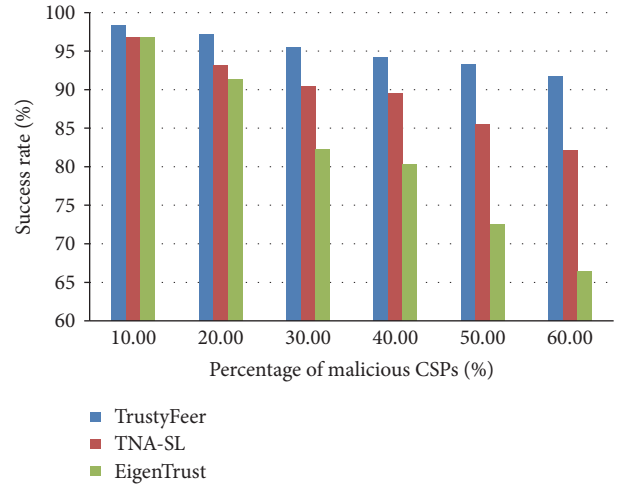


Figure 8: Success rate when different percentages of malicious CSPs are considered (collective).

Table 5: Experimental settings.

	Experiments	Number of CSPs	Number of malicious CSPs	Malicious model	Strategy
Set 1: different numbers of CSPs at a constant percentage of malicious CSPs	Ex 1.1	20	4 (20%)	Purely and feedback	Naïve/collective
	Ex 1.2	30	6 (20%)	Purely and feedback	Naïve/collective
	Ex 1.3	40	8 (20%)	Purely and feedback	Naïve/collective
	Ex 1.4	50	10 (20%)	Purely and feedback	Naïve/collective
	Ex 1.5	60	12 (20%)	Purely and feedback	Naïve/collective
	Ex 1.6	70	14 (20%)	Purely and feedback	Naïve/collective
Set 2: constant number of CSPs at different percentages of malicious CSPs	Ex 2.1	40	4 (10%)	Purely and feedback	Naïve/collective
	Ex 2.2	40	8 (20%)	Purely and feedback	Naïve/collective
	Ex 2.3	40	12 (30%)	Purely and feedback	Naïve/collective
	Ex 2.4	40	16 (40%)	Purely and feedback	Naïve/collective
	Ex 2.5	40	20 (50%)	Purely and feedback	Naïve/collective
	Ex 2.6	40	24 (60%)	Purely and feedback	Naïve/collective

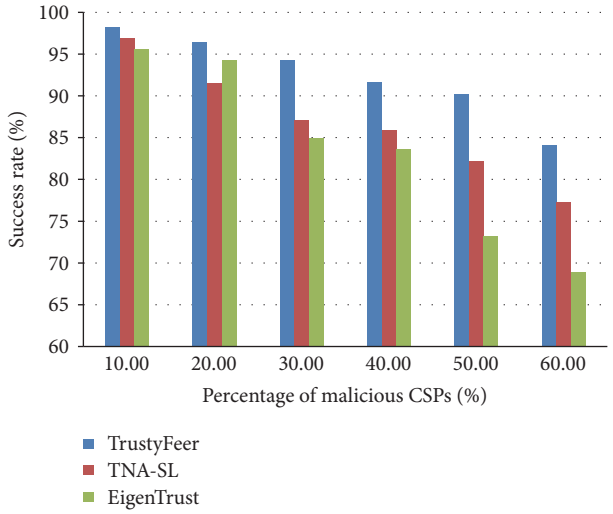


Figure 9: Success rate when different percentages of malicious CSPs are considered (naïve).

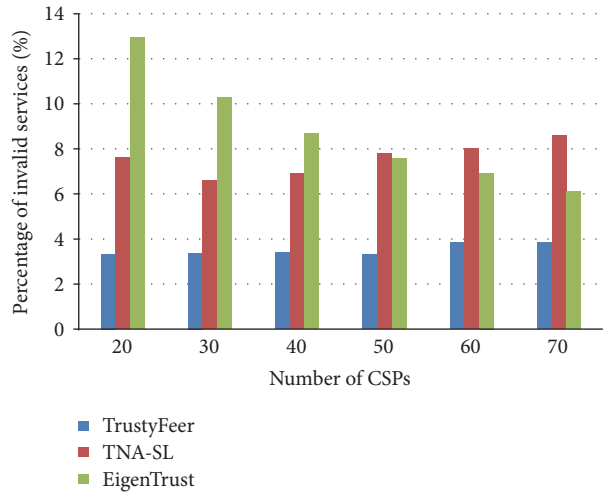


Figure 10: Percentage of invalid services when different numbers of CSPs are considered (collective).

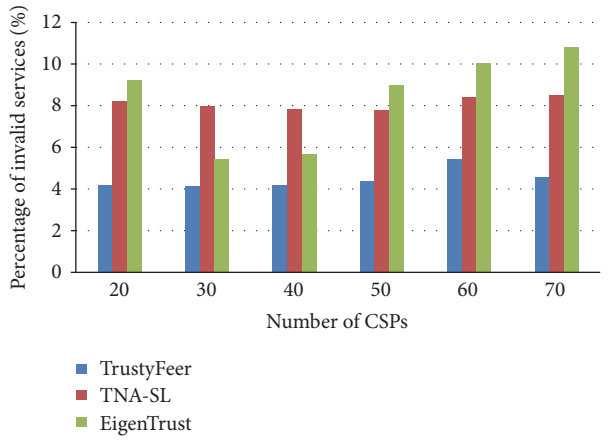


Figure 11: Percentage of invalid services when different numbers of CSPs are considered (naïve).

5.1. Success Rate. Figures 6–9 illustrate the success rates for good CSPs, calculated as the number of valid services received by good CSPs divided by the number of transactions attempted by good CSPs for each trust system.

In Figures 6 and 7, the success rate of good CSPs is plotted against the number of CSPs in the system. They clearly show that, at a constant percentage, 20%, of malicious CSPs, TrustyFeer and TNA-SL success rates remain almost constant as the number of CSPs increases. However, the TNA-SL success rate is much lower than that of TrustyFeer for all numbers of CSPs. This shows the scalability of TrustyFeer for any number of CSPs. In contrast, EigenTrust has an inconsistent success rate when the number of CSPs is considered. Comparing Figures 7 and 8 suggests that, in contrast to EigenTrust, TrustyFeer and TNA-SL are more powerful in handling complex collective malicious behaviors than naïve peers. Such ability has its root in the subjective logic robust feature where the trust value is calculated based on four factors rather than simply aggregating the rating from previous transitions which is the case with EigenTrust. The additional feature of weighting the feedback of each CSP by

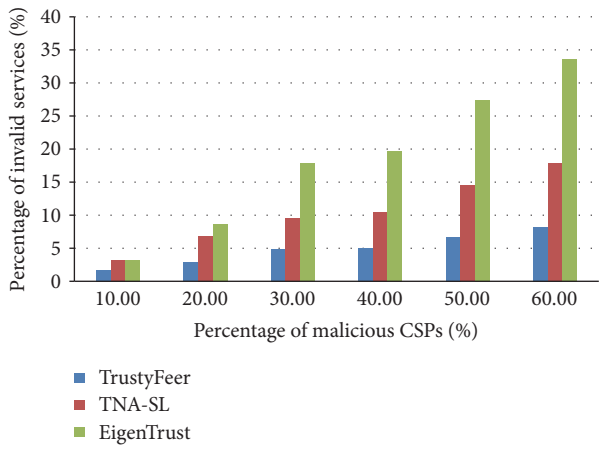


Figure 12: Percentage of invalid services when different percentages of malicious CSPs are considered (collective).

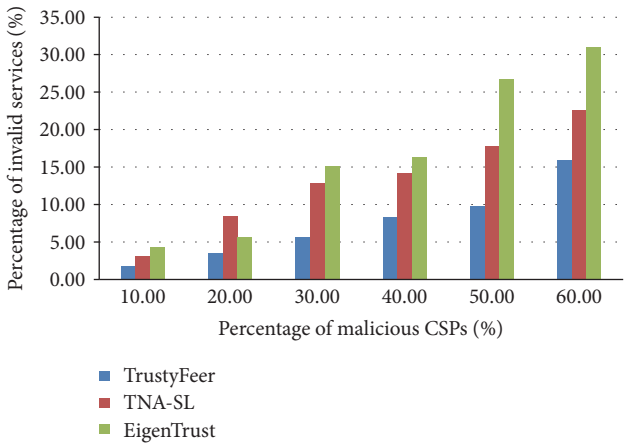


Figure 13: Percentage of invalid services when different percentages of malicious CSPs are considered (naïve).

its reputation value might be the main driving force behind the superior performance achieved by TrustyFeer.

Figures 8 and 9 plot the success rate of good CSPs against the percentage of malicious CSPs in the system, at a fixed number of 40 CSPs. Despite showing a slight decrease in the success rate as the percentage of malicious CSPs in the system increases, TrustyFeer still outperforms TNA-SL and EigenTrust in all scenarios. When the percentage of malicious CSPs is low, the difference is small. However, the gap increases as the percentage of malicious CSPs in the system increases. Such an observation might be strongly correlated with the way the opinion is calculated in TrustFeer, where the trust value of a CSP is weighted by its global reputation. This weighting strategy marginalizes the effect of malicious CSPs, especially of those forming collectives, as illustrated by Figure 8. In contrast, the centered strategy of the EigenTrust algorithm, around certain peers, proved to be inefficient when some of them turned out to be malicious which can easily happen in the malicious collectives model.

From the above discussion, we can clearly see that TrustyFeer surpasses TNA-SL and EigenTrust in helping good CSPs that conform to SLAs in exchanging services. This can be attributed to the fact that TrustyFeer uses a hybrid approach in calculating trust, while TNA-SL and EigenTrust use the peer-based approach.

5.2. Percentage of Invalid Services. In this set of results, the percentage of invalid services exchanged by good CSPs is evaluated for each reputation system.

Figures 10 and 11 illustrate the relationship between the percentage of invalid services exchanged by good CSPs and the number CSPs in the system. We can see that TrustyFeer outperforms TNA-SL and EigenTrust in terms of the percentage of invalid services exchanged by good CSPs. However, in all scenarios, the percentage of invalid services exchanged by good CSPs remains steady when TrustyFeer is used. Regardless of the P2P federated cloud scale, TrustyFeer has successfully identified the majority of malicious providers and isolated them, so, few good peers mistakenly exchange services with them. This behavior of TrustyFeer has its root in its weighting strategy where the opinion of any CSP is multiplied by its global reputation value. Additionally, the new proposed subjective logic formulas should also have a strong connection to this superior behavior as the SLA attribute values of previous transactions of each CSP are taken into consideration.

Figures 12 and 13 show the success of TrustyFeer in decreasing the percentage of invalid services compared to TNA-SL and EigenTrust. The percentage of invalid services increases steadily for all systems when the percentage of malicious CSPs increases. However, TrustyFeer manages to maintain its success despite such a large percentage of malicious CSPs. This is because TNA-SL and EigenTrust evaluate the quality of peers only when calculating trust values, as both are classified as peer-based reputation management systems. In contrast, being a hybrid system, TrustyFeer evaluates the quality of both the service and peers when calculating trust values, resulting in a more robust reputation management system. Two other factors can also be considered as driving this outstanding performance which are weighting peer's opinion by its reputation values and using new subjective logic formulas that take SLA parameters into consideration.

6. Conclusion

In P2P federated clouds, multiple CSPs can share their resources to facilitate the provision of efficient services to citizens in smart cities. However, associated with the interactions between these individual CSPs are several trust issues that affect the QoS of the federated cloud. In this paper, we have proposed TrustyFeer, a trust management system to enhance the QoS in P2P federated cloud environments. The TrustyFeer system proposes the use of subjective logic equations based on SLAs and the global reputation of CSPs to calculate trust values. The system was evaluated using a strictly controlled simulation environment with varying numbers of CSPs and threat models. In addition, the performance of TrustyFeer was compared against the TNA-SL and EigenTrust reputation management systems. The simulation results showed that TrustyFeer has a positive impact on reducing

the percentage of services that do not conform to SLAs and increasing the success rate of services that do conform to SLAs between good CSPs.

In future work, we plan to implement a P2P trust simulator especially for a P2P cloud environment. This will include a number of virtual machine requests between CSPs with different characteristics, based on a real data repository. This will help us to gain deep insights into system performance when evaluating a P2P cloud ecosystem.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Saudi Aramco, under the "Saudi Aramco Ibn Khaldun Fellowship for Saudi Women," in partnership with the Center for Clean Water and Clean Energy at MIT, and the Deanship of Scientific Research at King Saud University through Research Group no. RG-1438-002.

References

[1] M. J. Kaur and P. Maheshwari, "Building smart cities applications using IoT and cloud-based architectures," in *Proceedings of the International Conference on Industrial Informatics and Computer Systems, CIICS 2016*, UAE, March 2016.

[2] A. Zanella, N. Bui, A. P. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[3] P. Mell and T. Grance, "The NIST definition of cloud computing," *National Institute of Standards and Technology*, vol. 53, no. 6, 2009.

[4] T. H. Noor, Q. Z. Sheng, S. Zeadally, and J. Yu, "Trust management of services in cloud environments: obstacles and solutions," *ACM Computing Surveys*, vol. 46, no. 1, article 12, 2013.

[5] A. J. Ferrer, F. Hernández, J. Tordsson et al., "OPTIMIS: a holistic approach to cloud service provisioning," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 66–77, 2012.

[6] P. Riteau, "Building dynamic computing infrastructures over distributed clouds," in *Proceedings of the 1st IEEE Symposium on Network Cloud Computing and Applications, NCCA 2011*, pp. 127–130, fra, November 2011.

[7] M. R. M. Assis, L. F. Bittencourt, R. Tolosana-Calasanz, and C. A. Lee, "Cloud federations: requirements, properties, and architectures," in *Developing Interoperable and Federated Cloud Architecture*, Kecskemeti, G., A. Kertesz, and Z. Nemeth, Eds., Chapter 1, pp. 1–41, IGI Global, Hershey, Pa, USA, 2016.

[8] O. Shareef and A. Kayed, "A survey on federated clouds environment," *Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 2, pp. 83–92, 2015.

[9] O. Babaoglu, M. Marzolla, and M. Tamburini, "Design and implementation of a P2P Cloud system," in *Proceedings of the 27th Annual ACM Symposium on Applied Computing, SAC 2012*, pp. 412–417, Italy, March 2012.

[10] T. Hardjono, D. Shrier, and A. Pentland, *TRUST::DATA: A New Framework for Identity and Data Sharing*, Visionary Future LLC, 2016.

[11] U. Kaur and D. Singh, "Trust: models and architecture in cloud computing," *International Journal of Computer Science and Information Security (IJCSIS)*, vol. 13, no. 12, 150 pages, 2015.

[12] M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica, "Free-riding and whitewashing in peer-to-peer systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 1010–1018, 2006.

[13] F. Z. Filali and B. Yagoubi, "Global trust: a trust model for cloud service selection," *International Journal of Computer Network and Information Security*, vol. 7, no. 5, pp. 41–50, 2015.

[14] P. Resnick and R. Zeckhauser, "Trust among strangers in internet transactions: empirical analysis of eBay's reputation system," *Advances in Applied Microeconomics*, vol. 11, pp. 127–157, 2002.

[15] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks*, vol. 56, no. 18, pp. 3825–3833, 2012.

[16] J. Huang, F. Nie, H. Huang, and Y.-C. Tu, "Trust prediction via aggregating heterogeneous social networks," in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM 2012*, pp. 1774–1778, USA, November 2012.

[17] K. K. Bharadwaj and M. Y. H. Al-Shamri, "Fuzzy computational models for trust and reputation systems," *Electronic Commerce Research and Applications*, vol. 8, no. 1, pp. 37–47, 2009.

[18] N. Iltaf and A. Ghafoor, "A fuzzy based credibility evaluation of recommended trust in pervasive computing environment," in *Proceedings of the 2013 IEEE 10th Consumer Communications and Networking Conference, CCNC 2013*, pp. 617–620, USA, January 2013.

[19] S. Song, K. Hwang, and Y.-K. Kwok, "Risk-resilient heuristics and genetic algorithms for security-assured grid job scheduling," *IEEE Transactions on Computers*, vol. 55, no. 6, pp. 703–719, 2006.

[20] S. Song, K. Hwang, R. F. Zhou, and Y.-K. Kwok, "Trusted P2P transactions with fuzzy reputation aggregation," *IEEE Internet Computing*, vol. 9, no. 6, pp. 24–34, 2005.

[21] P. Sarathi Banerjee, J. Paulchoudhury, and S. R. Bhadra Chaudhuri, "Fuzzy membership function in a trust based AODV for MANET," *International Journal of Computer Network and Information Security*, vol. 5, no. 12, pp. 27–34, 2013.

[22] M. Tavakolifard and S. J. Knapskog, "A probabilistic reputation algorithm for decentralized multi-agent environments," *Electronic Notes in Theoretical Computer Science*, vol. 244, pp. 139–149, 2009.

[23] W. T. Teacy, M. Luck, A. Rogers, and N. R. Jennings, "An efficient and versatile approach to trust and reputation using hierarchical Bayesian modelling," *Artificial Intelligence*, vol. 193, pp. 149–185, 2012.

[24] Y. Zhang and Y. Fang, "A fine-grained reputation system for reliable service selection in peer-to-peer networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 8, pp. 1134–1145, 2007.

[25] T. Muller and P. Schweitzer, "On Beta Models with Trust Chains," in *Proceedings of the Trust Management VII: 7th IFIP WG 11.11 International Conference, IFIPTM*, Malaga, Spain, 2013.

[26] T. van Deursen, P. Koster, and M. Petković, "Hedaquin: a reputation-based health data quality indicator," *Electronic Notes in Theoretical Computer Science*, vol. 197, no. 2, pp. 159–167, 2008.

[27] A. Jøsang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279–311, 2001.

[28] A. Jøsang, R. Hayward, and S. Pope, "Trust network analysis with subjective logic," in *Proceedings of the 29th Australasian Computer Science Conference (ACSC2006), CRPIT*, Hobart, Australia, January 2006.

[29] R. Zhou and K. Hwang, "Trust overlay networks for global reputation aggregation in P2P grid computing," in *Proceedings of the 20th IEEE International Parallel Distributed Processing Symposium, Rhodes Island*, 2006.

[30] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The EigenTrust algorithm for reputation management in P2P networks," in *Proceedings of the 12th International Conference on World Wide Web (WWW '03)*, pp. 640–651, Budapest, Hungary, May 2003.

[31] J. Wang and J. Liu, "The comparison of distributed P2P trust models based on quantitative parameters in the file downloading scenarios," *Journal of Electrical and Computer Engineering*, vol. 2016, Article ID 4361719, 2016.

[32] Z. Su, L. Liu, M. Li, X. Fan, and Y. Zhou, "Reliable and resilient trust management in distributed service provision networks," *ACM Transactions on the Web (TWEB)*, vol. 9, no. 3, article no. 14, 2015.

[33] N. Alhadad, Y. Busnel, P. Serrano-Alvarado, and P. Lamarre, "Trust evaluation of a system for an activity with subjective logic," in *Proceedings of the International Conference on Trust, Privacy and Security in Digital Business*, vol. 8647, pp. 48–59, Munich, Germany, 2014.

[34] E. Choo, J. Jiang, and T. Yu, "COMPARS: toward an empirical approach for comparing the resilience of reputation systems," in *Proceedings of the 4th ACM Conference on Data and Application Security and Privacy, CODASPY 2014*, pp. 87–98, USA, March 2014.

[35] H. A. Kurdi, "HonestPeer: An enhanced EigenTrust algorithm for reputation management in P2P systems," *Journal of King Saud University—Computer and Information Sciences*, vol. 27, no. 3, pp. 315–322, 2015.

[36] "QTM: P2P Trust Simulator, 2009," https://rtg.cis.upenn.edu/qtm/p2psim.php3.

[37] L. Xiong and L. Liu, "PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 7, pp. 843–857, 2004.

[38] R. Zhou and K. Hwang, "PowerTrust: a robust and scalable reputation system for trusted peer-to-peer computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 4, pp. 460–473, 2007.

[39] R. Chen, X. Zhao, L. Tang, J. Hu, and Z. Chen, "CuboidTrust: a global reputation-based trust model in peer-to-peer networks," in *in Autonomic and Trusted Computing*, vol. 4610, pp. 203–215, Springer, Hong Kong, China, 2007.

[40] E. Ayday and F. Fekri, "BP-P2P: Belief propagation-based trust and reputation management for P2P networks," in *Proceedings of the 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON 2012*, pp. 578–586, Republic of Korea, June 2012.

[41] R. Zhou and K. Hwang, "Gossip-based reputation aggregation for unstructured peer-to-peer networks," in *Proceedings of the 21st International Parallel and Distributed Processing Symposium, IPDPS 2007*, USA, March 2007.

[42] L. Huaiqing, W. Xuezhi, and L. Haitao, "Hierarchical fuzzy trust management for peer-to-peer network," in *Proceedings of the 2009 Second ISECS International Colloquium on Computing, Communication, Control, and Management, CCCM 2009*, pp. 377–380, China, August 2009.

[43] H. Zhao and X. Li, "VectorTrust: Trust vector aggregation scheme for trust management in peer-to-peer networks," in *Proceedings of the 18th International Conference on Computer Communications and Networks (ICCCN '09)*, IEEE, August 2009.

[44] E. D. Canedo, R. Junior, and R. Albuquerque, "Trust model for reliable file exchange in cloud computing," *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 4, no. 1, 2012.

[45] N. Dladlu and O. O. Ekabua, "Implementation of a novel peer-to-peer reputation-based trust management model in a cloud service provisioning environment," in *The International Conference on Digital Information Processing, E-Business and Cloud Computing*, Kuala Lumpur, Malaysia, 2016.

[46] I. Petri, O. F. Rana, Y. Rezgui, and G. C. Silaghi, "Trust modelling and analysis in peer-to-peer clouds," *International Journal of Cloud Computing*, vol. 1, no. 2/3, p. 221, 2012.

[47] J. Abawajy, "Determining service trustworthiness in intercloud computing environments," in *Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms, and Networks, I-SPAN 2009*, pp. 784–788, Taiwan, December 2009.

[48] S. Y. Lee, O.-H. Kwon, J. Kim, and S. J. Hong, "A reputation management system in structured peer-to-peer networks," in *Proceedings of the 14th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE '05)*, pp. 362–367, June 2005.

[49] H. Kurdi, S. Alnasser, and M. Alhelal, "AuthenticPeer: a reputation management system for peer-to-peer wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 11, no. 11, Article ID 637831, 2015.

[50] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, 2007.

[51] S. Ruohomaa, L. Kutvonen, and E. Koutrouli, "Reputation management survey," in *Proceedings of the 2nd International Conference on Availability, Reliability and Security, ARES 2007*, pp. 103–111, Australia, April 2007.

[52] D. Kaur and J. S. Gupta, "Proposed P2P reputation-based model to secure grid," in *Proceedings of the International Conference on Recent Advances Trends in Information Technology (iRAFIT)*, Patiala, Punjab, India, 2012.

[53] F. G. Mármol and G. M. Pérez, "Security threats scenarios in trust and reputation models for distributed systems," *Computers & Security*, vol. 28, no. 7, pp. 545–556, 2009.

[54] B. Wickremasinghe, R. N. Calheiros, and R. Buyya, "CloudAnalyst: a cloudsim-based visual modeller for analysing cloud computing environments and applications," in *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA '10)*, pp. 446–452, Perth, Wash, USA, April 2010.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 9590576, 13 pages
https://doi.org/10.1155/2018/9590576

Research Article

Multicriteria Parent Selection Using Cognitive Radio for RPL in Smart Grid Network

Adisorn Kheaksong,[1] Kanabadee Srisomboon,[1] Akara Prayote,[2] and Wilaiporn Lee [1]

[1]*Communication and Computer Network Research Group (C2NRG), Electrical Engineering, Department of Electrical and Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*
[2]*Computer Science, Department of Computer and Information Science, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

Correspondence should be addressed to Wilaiporn Lee; wilaiporn.l@eng.kmutnb.ac.th

Received 22 August 2017; Revised 8 January 2018; Accepted 23 January 2018; Published 21 February 2018

Academic Editor: Danda B. Rawat

To maintain reliability of advanced metering infrastructure network in smart grid, data sent from a smart meter must reach a data concentrator unit efficiently. Parent selecting mechanism in routing protocol for low-power and lossy (RPL) is a key to maintain the reliability by balancing workload of meters in the network. In this paper, a parent selecting mechanism with three criteria including expected transmission count, residual energy, and expected transmission time is proposed to improve workload balancing and lifetime differences of all meters. A meter selects an immediate parent based on three factors. From simulation results, parents' workload is better balanced and the lifetime of all meters in the network is depleted nearly at the same time. Moreover, a simulation with cognitive radio enabled meters, where data can be transmitted on a licensed channel opportunistically when the channel is not utilized, shows an improvement in the packet delivery ratio.

1. Introduction

Depending on an increasing of electrical usage, an electrical grid where the grid is connected by high-voltage network to local distribution systems cannot meet the requirement of electrical demand. Smart grid [1–5] is defined as the next generation of electrical grid system which integrates several communication technologies and digital data processing to improve the electric system reliability and efficiency. An advanced metering infrastructure (AMI) [6–8] is the application that allows a two-way communication between in-home smart meters and utilities in order to exchange the information—energy usage, control message, and emergency notification alert—between utility and clients via local access point (data concentrator unit: DCU). Therefore, once a malfunction occurs, it is reported to utility within real-time via AMI network. Then, the utility is able to repair the system remotely or sends a maintenance personnel to repair the system timely. On the other hand, the customer benefits from an immediate electrical usage feedback where they can manage their behavior of electrical utilization in economical purposes.

To support several application of AMI, routing protocol for low-power and lossy (RPL) is standardized by Internet Engineering Task Force (IETF) and is expected to be exploited in practice [9–12]. RPL is defined as a routing protocol based on distance vector, where the data route—directed acyclic graphs (DAG)—is constructed according to routing metrics defined by objective function (OF) [13]. In the RPL protocol, there are two main issues that need to be concerned including the meter (or node) lifetime and the transmission reliability [14, 15].

For AMI network in smart grid, low-power smart meters are organized as a rooted tree architecture where DCU is the root. The meter lifetime is the critical issue that affects reliability of smart grid system. To maintain the system reliability, data from meters needs to reach the utility forwarded by DCU. There are two situations where the data from meters cannot be transmitted to DCU. First, the battery of the meter depletes. Once a battery of a meter is depleted, not only the

meter's data cannot be transmitted to the DCU, but also the data packet of its dependent meters (successors) cannot reach to DCU. Second, the battery of the meter's relay node (or immediate parent) depletes and there is no candidate parent to forward the data from the meter. Consequently, it is up to maintenance to keep monitoring individual meter to avoid battery depletion. Therefore, it has to be mentioned that when a battery of a meter depletes, the data of the meter and its dependent meters cannot reach the DCU. This is one of main issues of AMI networks known as a network partitioning [16].

To facilitate the maintenance job in a battery depletion avoiding, the lifetime of all meters should be nearly the same. Therefore, the workload of meters in the network needs to be balanced. Parent selection mechanism plays an important role in balancing the meters' workload. In a traditional RPL, the immediate parent is selected by a child meter (child node) and determined by the quality of transmission channel only. The quality of transmission channel is based on expected transmission count (ETX). When the quality of transmission channel is best, ETX becomes 1 as the minimum. Once the quality drops, the value of ETX increases. With this mechanism, the meters with the better quality of transmission channel (lower ETX) will be selected as immediate parents and their workload is higher than others so that their battery will deplete fast. While meters with worse quality of transmission channel do not have much workload, their battery will deplete slower. Once the meters' workload are not balanced, the partitioned network will occur when early depletion of high workload parent occurs.

To balance the meters' workload, the workload needs to be equally split by minimizing an influence of channel quality in a parent selection. Early parent selecting functions focus on either channel quality or energy consumption [17–19]. By selecting the parent on energy constrain only, the meters may suffer from efficient route. Later, an energy-efficient oriented algorithm (EERA) [20] is proposed to balance the workload of parent by determining both channel quality and parent's residual energy. However, the appropriate weights have not been investigated and the detected packet delivery ratio (PDR) has not been studied. Moreover, the expected transmission time (ETT) [21], one of the RPL performance metrics, has not been taken into account yet. It should be noticed that several literatures have not considered the ETT because the communication of RPL is based on only a single protocol—Zigbee. By considering only Zigbee protocol, several transmitting constraints are normally the same, that is, channel bandwidth, maximum bit rate, transmit power, and so on. Therefore, ETT has not been taken into account. Based on an emerging of cognitive radio technology, the available licensed band can be exploited. Therefore, RPL protocol can adopt the CR technology to improve the workload balancing and also the PDR.

CR [22–24] is the technology that allows a dynamic utilizing of the unused communication channels by reconfiguring its operating parameters and functionalities. By allowing an unlicensed user to operate on the unused channel, an issue of spectrum scarcity and data congestion can be resolved. To exploit the CR technology, the devices in the network have to comprise cognitive capabilities where changing surrounding spectrum environment can be evaluated in order to determine the appropriate transmission plans. The first standardized air interface for CR network—IEEE 802.22 [25, 26]—is based on an opportunistic utilization on the white spaces within the television bands between 54 and 862 MHz, especially within rural areas where usage may be lower.

The main contribution of this paper is to improve the workload balancing and lifetime differences of all meters. Therefore, we propose a new parent selection mechanism, multicriteria parent selection (MPS) algorithm, which balances the meter workload considered from three criteria. The three criteria are based on three main RPL performance factors—ETX, residual energy, and ETT. By analyzing the appropriate parent from multiobjective [27–30], the workload of parent meters can be balanced and the best route is selected. Moreover, with CR enabled meters, MPS shows an improvement in the PDR where data is transmitted on a licensed channel opportunistically when the channel is not utilized. As a result, workload balancing and lifetime differences of all meters as compared to traditional RPL mechanism and energy-efficient oriented algorithm (EERA) are improve. In addition, we also investigate the appropriate distance between meters which achieves an acceptable PDR rate where the rate is as high as 0.9.

The remainder of this paper is organized as follows. Section 2 gives brief introduction RPL protocol. In Section 3, the issues of RPL for smart grid network are described. In Section 4, the framework of MPS algorithm is described in detail. The simulation results are shown in Section 5. Finally, conclusions are presented in Section 6.

2. IPv6 Routing Protocol for Low-Power and Lossy (RPL)

RPL [9] is defined as a routing based distance vector protocol that provides an information distribution over a network topology which is constructed dynamically. Once the nodes in RPL topology operate autonomously, it can minimize some data configurations in the nodes by utilizing trickle timer [31]. To implement with AMI network, nodes in the network are constrained by processing power, memory, and energy. The key aspect of RPL protocol is to connect the nodes in the network using directed acyclic graphs (DAG) without any cycle presented. In a DAG, a rooted tree routing topology is constructed and is called destination-oriented directed acyclic graphs (DODAG).

As depicted in Figure 1, the structure of RPL network can be determined as dynamically constructed rooted tree topology where the lowest node transmits the data to the root node via the upper nodes. There are three types of node including gateway, router, and client. In the topology, root node is determined as a topology gateway. The duty of the rest nodes are determined by the distance from the root and are assigned by the level of rank. The level of rank is assigned to each node in the network related to the nodes' position with respect to the root. If a node is far from the root node, a rank level of the node is high. On the other hand, if a node is placed near to the root node, the level of rank is low. Therefore, the level of rank increases in a downward

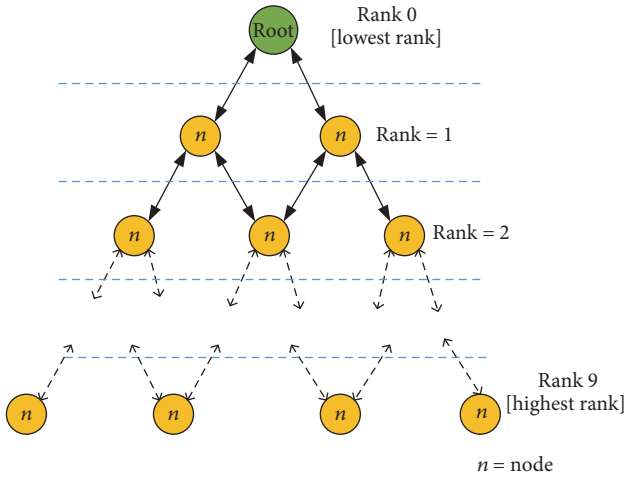


Figure 1: Structure of RPL network.

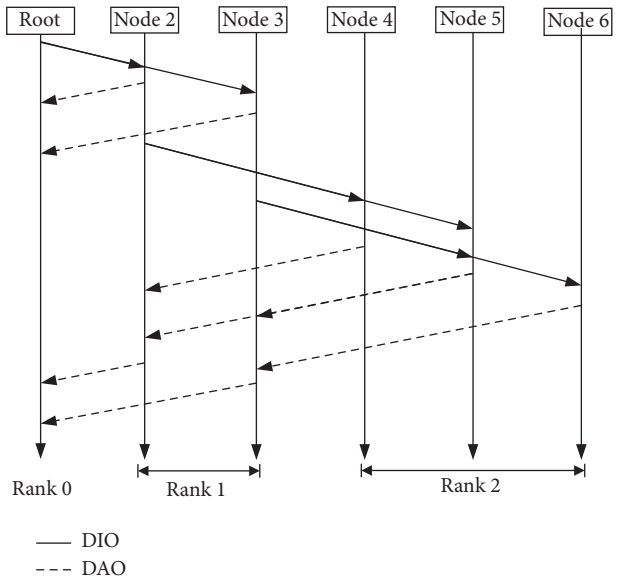


Figure 2: Route construction of RPL.

direction with respect to the root node. The node at the lower level of rank services a node at the higher level of rank as a parent. Therefore, the latest level of rank can transmit the data packet toward the root node via several ancestors (lower rank nodes). To prevent a cycle presented, the node in the topology is regulated to associate with only the same or lower level of ranks.

To construct the DAG, three control messages—DAG information option (DIO), destination advertisement object (DAO), and DODAG information solicitation (DIS)—are utilized and the routing metrics are defined by objective function (OF) [13]. First, as depicted in Figure 2, the root node broadcasts a DIO message which contain the information—rank of broadcasting node, OF, and DADAG-ID—to other nodes in the network. Nodes, which receive DIO message and desire to join the DAG, determine their own rank level using the received DIO message. Moreover,

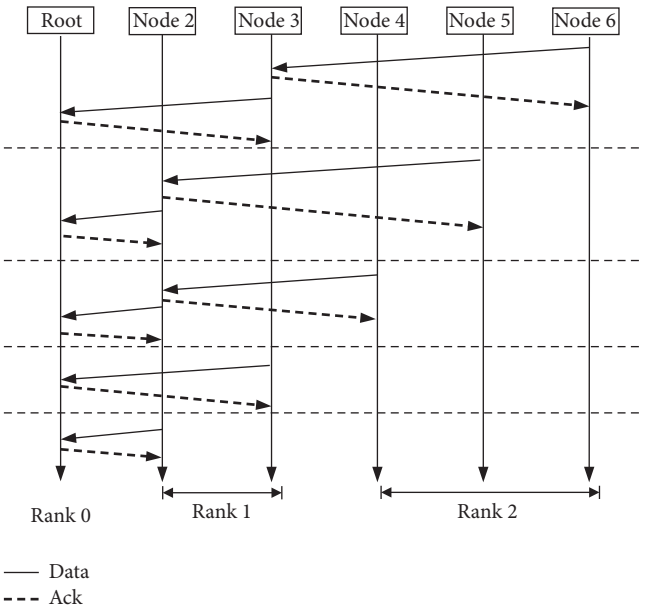


Figure 3: Data transmission period of RPL.

the node which transmitted the DIO message is added to the candidate parent list of receiver node. Once the receiver node receives more than one DIO message, the parent is selected from the candidate parent list constrained by objective metrics in OF. The best transmission channel serviced by the best parent is selected through objective metrics defined by OF. Therefore, a system reliability can be maintained. It should be noted that if the rank of receiver node is not the latest level of rank, it has to update its rank on DIO message and forward it to its neighbors.

To complete the route construction in an outward direction (root to leaf node), DAO message is transmitted in a unicast manner in an upward direction, as shown in Figure 3. In addition, the node will record the information of reverse path. After DAG is completely constructed, the higher level rank nodes are allowed to transmit the data packet in an inward direction—from node to root—via several appropriate ancestors nodes.

In RPL on AMI network, there are three communication schemes—route construction, data transmission, and notification alert transmission. First, DIO and DAO messages are transmitted to construct and maintain the RPL network. Second, based on AMI framework, the energy usage is transmitted from the leaf nodes to the DCU (root node) via their immediate ancestors (a number of parents). In general, the meter has to send the information about energy usage every 15 minutes to the DCU for real-time processing of smart grid technology. However, in this simulation, we set that the meters send the information to the DCU every 5 minutes for evaluating the routing performance. In an RPL protocol, once meters in DODAG have routing entry, the smart meter at the highest rank can send data traffic to the DCU toward immediate parent. Then, the parent will transmit the received data to his parent. It means that the data packet from leaf meter is transmitted to the DCU in

a hop-by-hop fashion via its ancestors of downward. Third, the notification alert message is transmitted to the DCU which needs to be performed in real-time. In this paper, we focus on route construction and data transmission in order to evaluate the performance of parent selection function. It should be noted that the size of data packet during energy usage transmission interval is larger than route construction; therefore, the size of data packet affects the PDR where the PDR may decrease. Therefore, the PDR of route construction and data transmission has to be investigated individually.

Traditionally, RPL mechanism defines only a single routing metric—the expected transmission count (ETX)—as a constraint for parent selection. ETX is the number of transmissions by a meter and is expected to be received by the destination. Then, the ETX can be expressed as

$$\text{ETX} = \frac{1}{\left(D_f * D_r\right)} = \frac{1}{1-p}, \tag{1}$$

where D_f is the measured probability that a packet is received by the neighbor, D_r is the measured probability that the acknowledgment packet is successfully received, and p is the error rate of data transmission.

In general, the parent is selected by determining ETX of the path where desired value of ETX is as low as 1 that is referred to a high channel quality. To determine the transmission path of each node, an objective function—Minimum Rank with Hysteresis Objective Function (MRHOF) [17]—is considered. To select the smallest path cost, there are two procedures which need to be considered. First, the path with minimum cost is determined. Second, the node selects the minimum cost path if the cost is better than the current path determined by a given threshold. The second procedure is called "hysteresis." Therefore, the route with the smallest path cost is selected. For achieving the objective of RPL, MRHOF can be used with several performance metrics—energy consumption, latency, or ETX—for selecting the best path quality.

As mentioned earlier, by selecting a parent use of a single routing metric cannot achieve the requirements of RPL where the node (meter) lifetime needs to be prolonged while the transmission reliability determined by packet delivery ratio needs to be maintained and stabilized. Therefore, an energy-efficient oriented algorithm (EERA) [20] combines these two objectives metrics—ETX and energy—to balance the workload of parent node while selecting the best route. To select the parent, a weight of importance between energy consumption and channel quality is considered. Therefore, a routing score for parent selecting is given by

$$\begin{aligned} R &= \alpha \frac{\text{ETX}}{\text{Maximum ETX}} \\ &\quad + (1-\alpha)\left(1 - \frac{\text{Remaining Energy}}{\text{Maximum Energy}}\right), \end{aligned} \tag{2}$$

$$\text{Remaining energy} = \text{Battery} - \text{Energy Consumption}$$

where α is the weight of importance between energy and ETX and the term of ETX is an intermediate ETX of a route between candidate parent and node. Once the energy consumption cannot be gathered directly from simulation, it can be calculated by the following equation:

$$\begin{aligned} \text{Energy Consumption} = {} & T_{\text{CPU}} * P_{\text{CPU}} + T_{\text{RadioRX}} \\ & * P_{\text{RadioRX}} + T_{\text{RadioTX}} \\ & * P_{\text{RadioTX}}, \end{aligned} \tag{3}$$

where T_{CPU} is the processing time of CPU, P_{CPU} is the power consumed by CPU, T_{RadioRX} is the processing time of the receiver, P_{RadioRX} is the power consumed by receiver, T_{RadioTX} is the processing time of the transmitter, and P_{RadioTX} is the power consumed by transmitter.

3. Problem Statement

In this section, we describe the issues that affect the performance of RPL protocol of AMI network including meter lifetime and transmission reliability.

3.1. Meter Lifetime. In this subsection, we describe the effect of the nonbalancing of energy. Based on a binary tree structure of AMI network, the root node acts as a DCU responsible for collecting the data from other nodes and also constructing the RPL process. The rest of nodes (or the node in the other ranks) act as an in-home smart meter.

As shown in Figure 4(a), during data transmission process, each node forwards the data to the root node via its immediate parents where the immediate parent is chosen in the route maintenance process. In this case, nodes 4 and 5 choose node 2 as their immediate parent and node 6 chooses node 3 to be its parent. As the passing of time, the battery of node 2 is consumed faster than node 3. Then, an early depletion of node occurs when node 2 is disabled. After node 2 depletes, node 4 and node 5 broadcast DIS message to request a route maintenance. As shown in Figure 4(b), node 3 receives the DIS message from node 5 but the DIS message from node 4 cannot reach node 3 and the root. Therefore, the communication of node 5 and node 6 are still performed while node 4 is terminated. Then, it results in a network partitioning.

Figure 5 shows a diagram of data transmission of node 4, node 5, and node 6 when node 2 depletes. As shown in Figure 5(a), when node 2 depletes and node 3 still communicates the root node, node 4 broadcasts DIS message to neighbors. After the DIS message is transmitted, node 4 waits for an acknowledgement in a period of waiting time. After the waiting time is out, node 4 will repeat broadcasting the DIS message and waiting for acknowledgement from neighbors until its battery runs out. It can be seen that not only the data of node 4 cannot reach the root node, but also it is a waste of energy for broadcasting DIS message. On the other hand, as shown in Figures 5(b) and 5(c), the data transmission of node 5 and node 6 still reaches the root node via node 3 because node 5 can complete a route maintenance process.

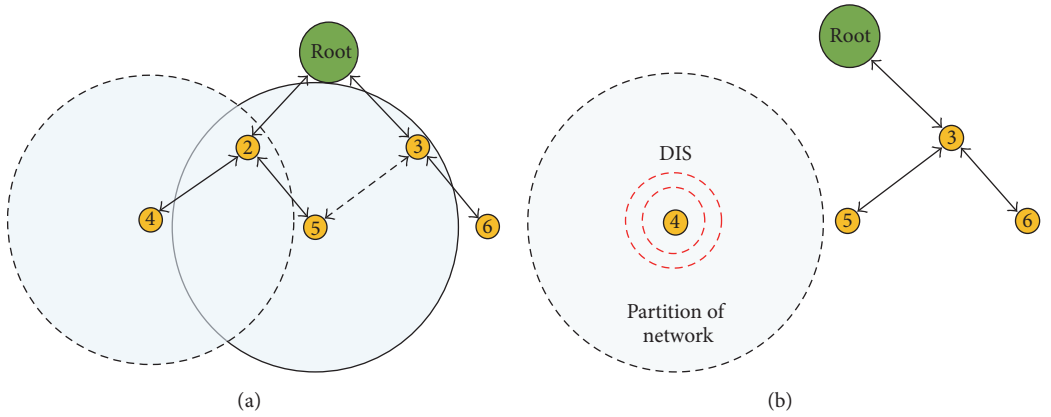


FIGURE 4: Occurrence of network partitioning, (a) network architecture of 3 ranks, and (b) network architecture of 3 ranks when node 2 is disabled.

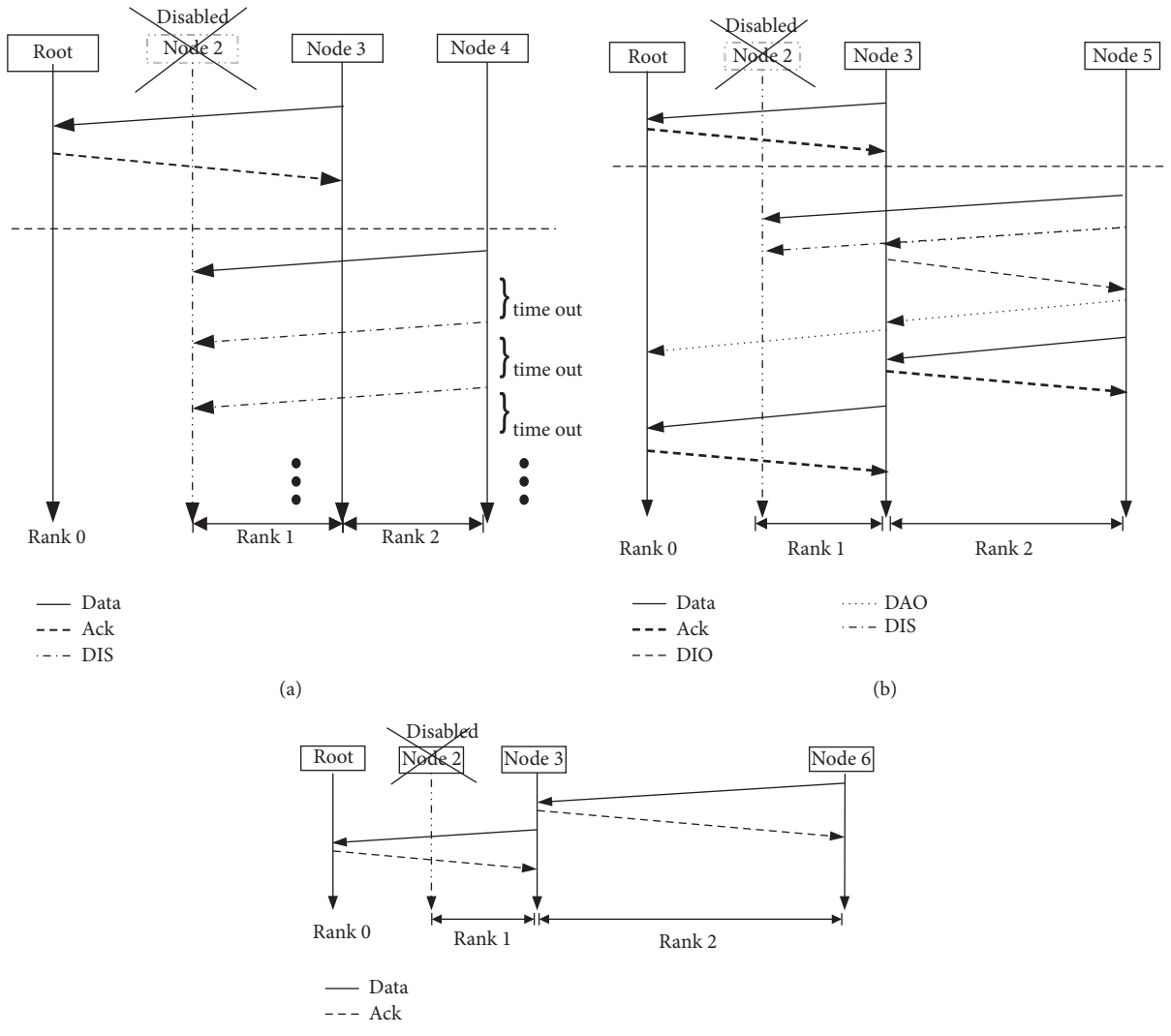


FIGURE 5: Data transmission framework of node when node 2 is disabled; (a) node 4, (b) node 5, and (c) node 6.

Therefore, it is important to balance the workload of the nodes in the network to avoid an early depletion of node and network partitioning. Therefore, the lifetime of the lowest rank nodes (nodes 2 and 3) is very important for this scenario.

3.2. Transmission Reliability. In a smart grid network, even if smart meters are static, the environment such as noise, interference, and fading may affect the transmission reliability. In general, the transmission reliability is affected by two factors including node density and the quality of the transmission channel. In case of node density, when the number of nodes in the network is high, the nodes may transmit data packet at the same time and the transmission from a node may interfere with the transmission of other nodes. Then, the data packet may drop during transmission and cannot reach the parent node. Therefore, packet delivery ratio (PDR) decreases due to an increasing in the number of nodes in the network.

On the other hand, a quality of transmission channel is the main factor that affects the transmission reliability. In general, the channel quality can be determined by a ratio between received signal and noise power (SNR). In practice, the received signal power drops when the distance between nodes increases. Therefore, when a child node is far away from its parent, the SNR is poor. By this factor, the quality of received signal is worse than to be utilized and it is ignored by the parent which is referred to a decreasing in a packet delivery ratio (PDR). Then, the parent needs to request data retransmission. Since the meters have more load of data transmission, the energy consumption is higher so the battery depletes faster than expected. Therefore, the network partitioning occurs.

As mentioned earlier, we focus on the partitioned network issue; therefore, we set the network topology as a binary rooted tree architecture where the node in the same rank cannot communicate with the others due to being out of the transmission range. This kind of topology is important in practice and occurs in the rural area where each meter is far away from others; then it may communicate with its parents only. If the battery of its parent depletes, the data from the meter cannot reach the DCU. Since the meters are far away from others, the node density does not influence the transmission channel quality. Consequently, in this paper, the distance between nodes is the main factor that degrades the channel quality.

4. Multicriteria Parent Selection Algorithm

In this paper, we propose the multicriteria parent selection (MPS) algorithm to address two main problems—meter lifetime and stability of channel quality. Moreover, the expected transmission time (ETT) during energy usage data transmission is taken into account and used as a performance metric for selecting the immediate parent. To select the most appropriate parent, an analytic hierarchy process (AHP) is exploited by determining three critical parameters—expected transmission count (ETX), residual energy, and expected transmission time (ETT)—with assigned weight of importance. Moreover, to improve a PDR and increase an ability of workload balancing, a concept of cognitive radio (CR) is adopted. By utilizing concept of CR, smart meters are determined as CR enabled nodes where the meter must be implemented with CR functions. Once the bit-rate of CR channel and the transmission power are higher than Zigbee, the range of data transmission range under an acceptable PDR can be improved.

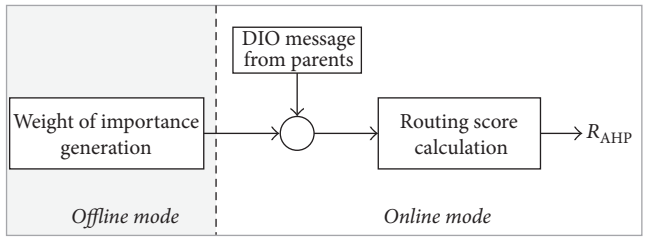


Figure 6: Operation of MPS mechanism.

As depicted in Figure 6, the operation of MPS can be separated into an offline and online mode. In an offline mode, the weight of importance between the main criteria—ETX, residual energy, and ETT—is determined. This can be done by performing a pairwise comparison between these main criteria. Then, the weights are generated. In an online mode, MPS gathers the data—values of ETX, residual energy, and ETT—from the current and candidate parents in order to select an immediate parent. In a subcriteria phase, data from parents is compared pairwise. The results from subcriteria pairwise comparisons then performed a summation according to the predetermined weight of importance in order to compute the routing score (R_{AHP}) of each parent. Finally, the routing score of current and candidate parents is compared to determine the immediate parent.

Then, the framework of the multicriteria parent selection (MPS) algorithm is categorized into two sections—parent selection during route construction and parent selection during energy usage data transmission.

4.1. Parent Selection during Route Construction. As stated by RPL provision mechanism, a control message is broadcast over network continually in order to form a network topology dynamically. As described in the previous section, once a network topology is formed, both parent node and candidate node send a DIO message to higher rank nodes where the higher rank nodes are determined as children. After a child receives a DIO message, it will determine which node is appropriate to be its parent. As mentioned earlier, to select the parent, route performance—ETX—cannot be merely determined. Therefore, other metrics—residual energy and ETT—need to be considered.

In this subsection, a routing selection using AHP algorithm is described. AHP is a powerful tool that combines several indicators performance into a single key performance indicator (or performance metrics) [32, 33]. Therefore, different weight of importance between performance indicators should be assigned to those performance indicators. By determining the weight using AHP, a pair comparison between indicators is determined as the weight. Therefore, an inconsistency of parent selecting can be addressed. To

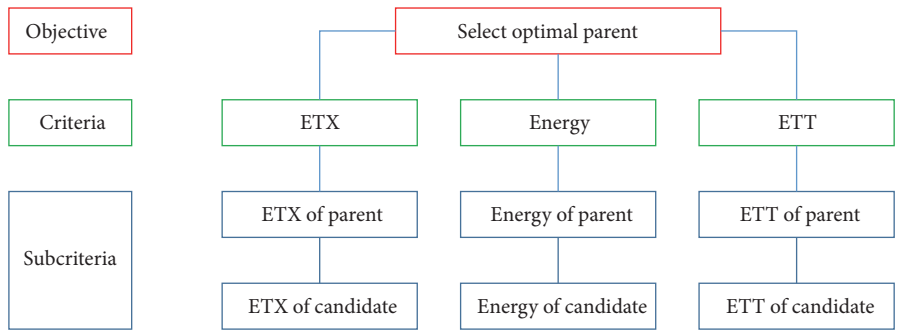


Figure 7: Structure of objective, criteria, and subcriteria.

determine the most appropriate parent while maintaining overall performance, ETX, residual energy, and ETT are determined as the parent selection criteria of AHP.

Then, the routing score of each parent can be determined as the following procedures:

(1) Define the objective. In this paper, the objective is to determine the most appropriate parent when ETX, residual energy and ETT are determined as the criteria. The subcriteria are the real value of ETX in each route, residual energy of parent, and the ETT of the route.

(2) Group the criteria, subcriteria, and alternatives. As described in the previous procedure, the structure of objective, criteria, and subcriteria is shown in Figure 7.

(3) Make a pairwise comparison of elements in each group.

(3.1) Construct $n \times n$ of pairwise comparison matrix ($\mathbf{A}$), where n is the number criteria, by comparing each criteria with respect to others. Then, the pairwise comparison matrix ($\mathbf{A}$) is shown in Figure 8 and can expressed as

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{bmatrix}, \quad (4)$$

where a_{12} represents the weight of importance between ETX and remaining energy, a_{13} represents the weight of importance between ETX and ETT, and a_{23} represents the weight of importance between remaining energy and ETT.

It should be noted that a_{ji} is equal to $1/a_{ij}$ when i and j are $1, 2, \ldots, n$. To compare the weight of important between criteria, the scale of range 9 to 1/9 is determined. For example, if a_{12} is 5, ETX is 5 times more important than energy. On the other hand, if a_{12} is 1/5, ETX is 5 times less important than energy.

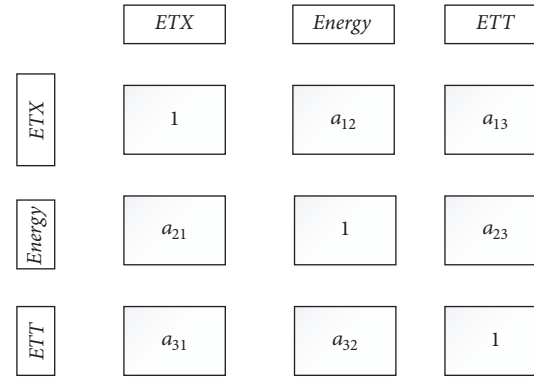


Figure 8: Pairwise comparison matrix ($\mathbf{A}$) of main criteria.

(3.2) Compute the normalize eigenvector. Then, the weight of importance of each criteria ($\mathbf{c} = [c_1, c_2, \ldots, c_n]$) is obtained.

(3.3) Do procedures (3.1) and (3.2) for subcriteria of each criteria. It should be noted that the weight of importance of subcriteria can be determined by the ratio of real values of compared elements.

(4) Calculate the weight of subcriteria according to the main criteria.

(5) Evaluate the alternatives according to the weight by summarizing all of calculated weights for each element.

4.2. Parent Selection during Energy Usage Data Transmission. In an AMI network, the node (meter) has to send the information about energy usage every 15 minutes to the DCU for real-time processing of smart grid technology. In this subsection, we propose a CR hopping algorithm where the nodes determine the worthiness of exploiting available channel of CR network when ETX, ETT, and residual energy are determined. As illustrated in Algorithm 1, the parameters—residual energy, ETX, and ETT—of immediate parent and candidate parent are used as the inputs of the algorithm. It should be noted that subscripts "im" and "c" are an immediate parent and candidate parent, respectively. P_i denotes the serviced parent. CH_i denotes the transmission

```
Input: E_im and E_c
Output: Parent and transmission channel
if E_im <= 0 and E_c <= 0 then
   return P_i ← 0;
else
  if E_im > 0 then
     Calculate R_AHP,Z of immediate parent using Zigbee channel;
     if E_c > 0 then
        Calculate R_AHP,C of candidate using CR channel;
     else
        Calculate R_AHP,im of candidate using CR channel;
  else
       if E_c > 0 then
          Calculate R_AHP,C of candidate using CR channel;
          Calculate R_AHP,im of candidate using Zigbee channel;
if E_im > 0 then
  if E_c > 0 then
     if R_AHP,Z > R_AHP,C then
        return P_i ← 1; CH_i ← 1; //
     else
        return P_i ← 2; CH_i ← 2; //
  else
     if R_AHP,Z > R_AHP,C then
        return P_i ← 1; CH_i ← 1; //
     else
        return P_i ← 1; CH_i ← 2; //
else
  if E_c > 0 then
     P_i ← 2; //
     if R_AHP,Z > R_AHP,C then
        return CH_i ← 1;//
     else
        return CH_i ← 2; //
  else
        return P_i ← 0; CH_i ← 0; //
```

Algorithm 1: CR Hopping during importance data transmission.

protocol where Zigbee and CR are determined. R_{AHP} is a calculated routing score of parent and candidate parent.

First, residual energy of immediate and candidate parent is determined. If the immediate and candidate parent deplete, the algorithm states that there is no enabled parent. Otherwise, the residual energy of immediate and candidate parent is compared to assign their transmission technology. If both immediate and candidate parent are enabled, the Zigbee and CR are assigned to immediate and candidate parent, respectively. If energy of either immediate or candidate parent is enabled, Zigbee and CR are assigned to the enabled parent, to calculate the routing score. Therefore, the routing score of enabled parent with Zigbee and routing score of enabled parent with CR technology is determined by $R_{\mathrm{AHP},Z}$ and $R_{\mathrm{AHP},C}$, respectively. Second, after transmission technology is assigned to both immediate and candidate parent, the routing score is calculated by using AHP algorithm as described in the previous subsection. Then, the routing scores are compared. The output of the algorithm declares two parameters including selected parent and the transmission technology where the interpretation of subscripted number of (i) is shown in Table 1.

In an information interaction overhead, in RPL protocol, the DIO message for route construction and maintenance is exchanged between nodes periodically. Our mechanism also follows this protocol; it does not incur any extra overhead. To select a parent in the route maintenance, the main computation of our mechanism is on the pairwise comparison of subcriteria, where an eigendecomposition theory is utilized. The process is bound at $O(n^3)$ where n is number of candidate parents.

5. Simulation Results

In this section, the performance of the multicriteria parent selection (MPS) algorithm is compared to two early parent selecting functions—traditional RPL and energy-efficient oriented algorithm (EERA) [20]. As mentioned earlier, the main objective of RPL protocol on smart grid technology is to distribute an information over a network topology while maintaining a system reliability using low-power devices and prolonging the meter lifetime. Therefore, the two vital performance metrics that are used to evaluate the network

Table 1: Interpretation of subscripted number.

Subscripted number (i)	0	1	2
CH_i	No enabled technology	Zigbee	CR
P_i	No enabled parent	Immediate parent	Candidate parent

Table 2: Simulated parameters.

Parameters	Value
Tx power	66.16 mW
Rx power	70.695 mW
Tx power in CR mode	198.51 mW
Idle power	900 nW
Sleep power	400 nW
Spectrum sensing power	65.83 mW
DIO interval before data transmitting	5 seconds
Data transmitting interval	5 mins
Data packet	127 bytes
DIO message size	64 bytes
DAO message size	46 bytes
DIS message size	2 bytes
Transmission time of data packet	192 us
Transition time from sleep to active mode	970 us

performance are packet delivery ratio (PDR) and the meter lifetime of the network.

5.1. Simulation Setup. In this paper, the two communication schemes—route construction and energy usage data communication—are determined. It should be noted that the notification alert and the performance of ETT will be considered in the future work. All simulation runs on MATLAB where the simulation time is 3600 minutes. Parameters which are set in the simulation are shown in Table 2.

As shown in Figure 1, the network topology is set as a binary rooted tree architecture and defined that the quality of left transmission channel is better than the right one defined by the distance between two meters. For example, for 3 ranks, the distance from node 5 to node 2 is less than node 5 to node 3 as depicted in Figure 9. In the scenario, the node in the same rank cannot communicate with the others due to the transmission range. Then, the consumption energy can be calculated by using (3). Moreover, we also investigate the transmission quality due to an increasing of transmission range. Then, the distance between nodes ranged between 10 and 100 meters.

Since one ancestor depletes, it means that the data transmission of children is terminated. Therefore, the meter lifetime is measured only when the data transmission can reach to the DCU. To evaluate the meter lifetime, the maximum, average, and minimum meter lifetime as a function of rank numbers are illustrated where the maximum, average, and minimum lifetime are obtained from all meters. The meter lifetime is determined by the time that the communication of a meter cannot reach the DCU.

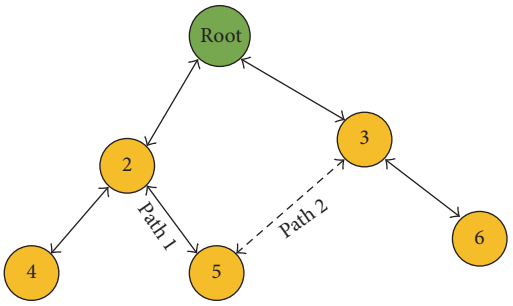


Figure 9: 3-rank network topology.

Workload balancing ability is considered through a relative balancing factor (RBF) which is given as

$$\mathrm{BF}_i = \left|\mathrm{Lifetime}_{\max} - \mathrm{Lifetime}_{\min}\right| \tag{5}$$

$$\mathrm{RBF} = \frac{\left|\mathrm{BF}_{\mathrm{RPL}} - \mathrm{BF}_i\right|}{\mathrm{BF}_{\mathrm{RPL}}}, \tag{6}$$

where BF is a balancing factor in each parent selection function, $\mathrm{Lifetime}_{\max}$ is the maximum lifetime of node in each rank, $\mathrm{Lifetime}_{\min}$ is the minimum lifetime of node in each rank, and i is the EERA or MPS.

To compare the improvement of partitioned network avoiding, the minimum meter lifetime of later parent selecting function is compared with respect to the minimum meter lifetime of traditional RPL. Then, the percentage of improvement (PM) can be expressed as

$$\mathrm{PM} = \left(\frac{\left|\mathrm{Lifetime}_{\mathrm{min,RPL}} - \mathrm{Lifetime}_{\mathrm{min},i}\right|}{\mathrm{Lifetime}_{\mathrm{min,RPL}}}\right) \times 100, \tag{7}$$

where $\mathrm{Lifetime}_{\min}$ is the minimum lifetime of node in each network and i is the EERA or MPS.

In order to compute the weights, every two factors performed pairwise comparison where the important weight between two factors need to be assigned. In this paper, we investigate the appropriate value of important weights and those values are set as follows: the importance between ETX and residual energy is equal and the ETX and residual energy are 2 times more important than ETT. Then, the pairwise comparison matrix ($\mathbf{A}$) is given by

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}. \tag{8}$$

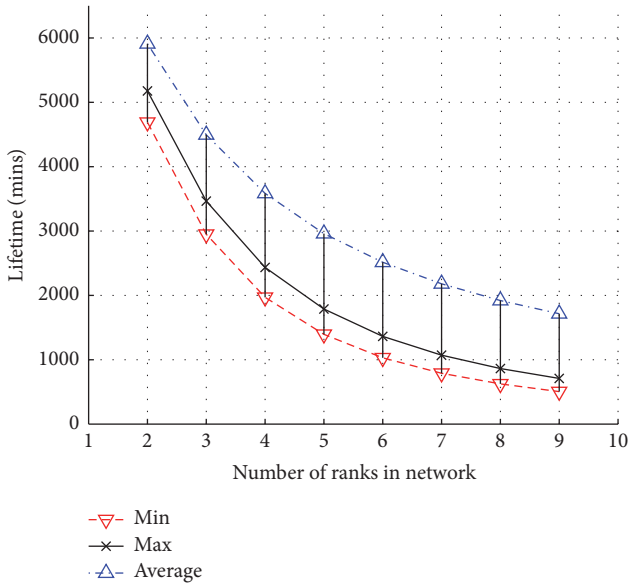


Figure 10: Node operation time of traditional RPL mechanism.

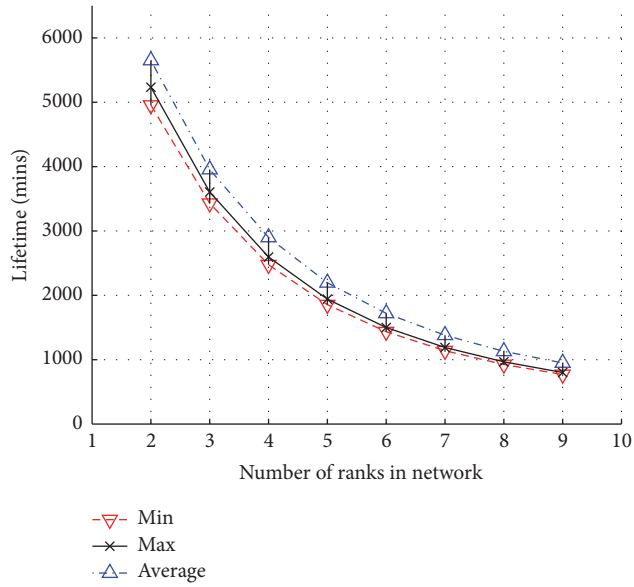


Figure 11: Node operation time of EERA.

5.2. Simulation Results. In this subsection, we first evaluate the performance of three parent selection functions—traditional RPL (RPL), EERA, and MPS—under different number of ranks.

First, the meter lifetime of each parent selection function is illustrated. As shown in Figure 10, by using traditional RPL (RPL), the maximum meter lifetime is much different from the minimum. Because the child meters keep forwarding the data to the parent who is nearer to it, for example, node 5 forwards the data to node 2 until the battery of node 2 depletes. After node 2 depletes, the data from node 4 cannot be transmitted to the DCU. Then, node 4 is partitioned from the network. On the other hand, before node 2 depletes, node 3 has only one child, node 6. After node 2 depletes, node 3 is determined as a relay of nodes 5 and 6. Therefore, the average meter lifetime of traditional RPL decreases when the number of ranks increases. As shown in Figure 11, since EERA selects the immediate parent by considering residual energy of both candidate parents and ETX. Therefore, the workload of parents is balanced since the maximum meter lifetime is near to the minimum. Moreover, it can be seen that the average meter lifetime increases as compared to the traditional RPL.

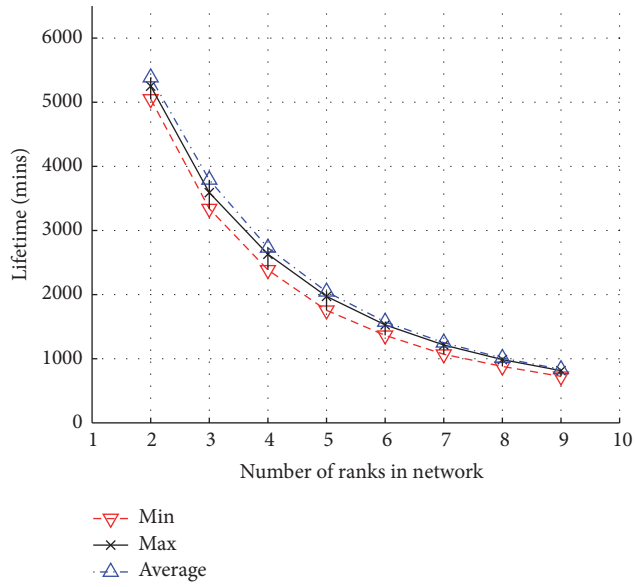


Figure 12: Node operation time of MPS.

As shown in Figure 12, for MPS algorithm, it can be seen that the maximum meter lifetime of MPS is nearly the same as the minimum which means that MPS balance the meter workload in the network efficiently. This is because MPS considers the three performance metrics—ETX, residual energy, and ETT—as the multiobjective and performing a pairwise comparison between them. Then, the appropriate parent is analyzed comprehensively. During data transmission, even if the CR mode is activated and it consumes more energy than Zigbee, MPS can still balance the meter workload in network.

As shown in Figure 13, the improvement of balancing ability of MPS and EERA is compared to traditional RPL through a relative balancing factor which can be expressed in (6). As the number of ranks increases, the number of children of the highest ancestors (nodes 2 and 3) increases. Then, the highest ancestors consumes more energy, once they have to pass on the children data to the root node. Since traditional RPL ignores the residual energy of the lower rank nodes to determine the immediate parent, the workload of parent nodes is not balanced. Therefore, when MPS and EERA take residual energy into account, they give better balancing ability than traditional RPL. As compared to EERA, MPS gives higher improvement of balancing ability for all ranks because MPS improves the balancing ability by considering

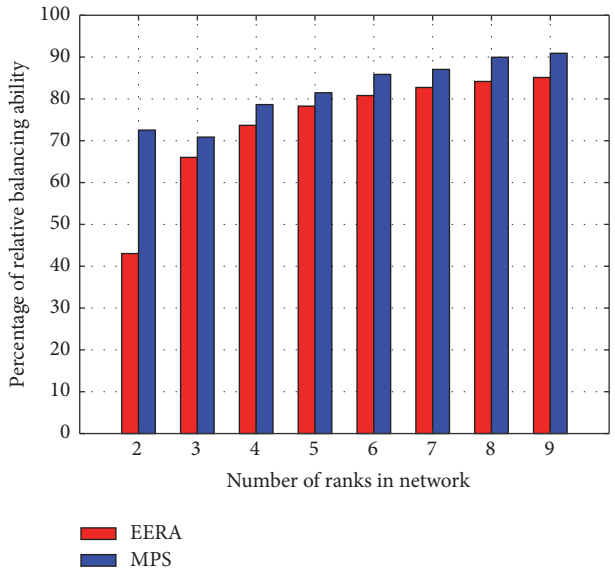


Figure 13: A balancing factors of MPS as compared to traditional RPL and EERA.

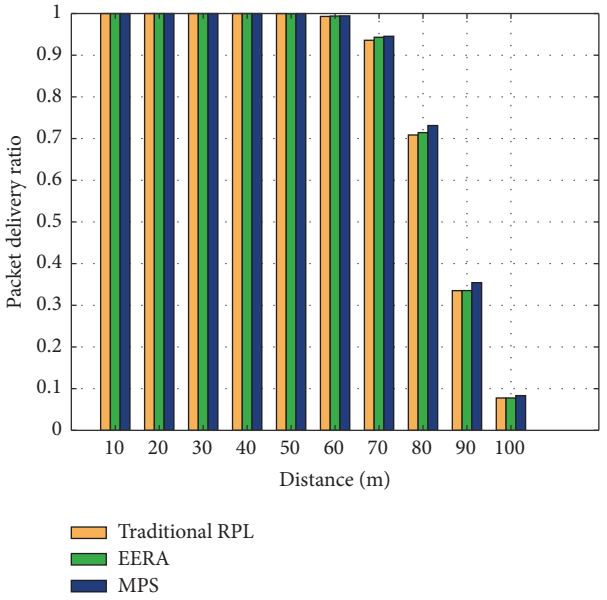


Figure 14: Average PDR of MPS as compared to RPL and EERA during route construction.

the routing factors more comprehensively and adopting CR channel.

Second, we evaluate the transmission reliability by comparing the average PDR as a function of distances. As shown in Figure 14, an average PDR decreases due to an increase in distance between nodes. It should be noted that, during route construction and maintenance, MPS does not activate CR mode for battery saving purpose. By using analyzing weights using AHP, not only the workload of parents is balanced, but also MPS gives the highest averaged PDR for all distances.

Third, we evaluate the transmission reliability during energy usage data transmission by comparing the average

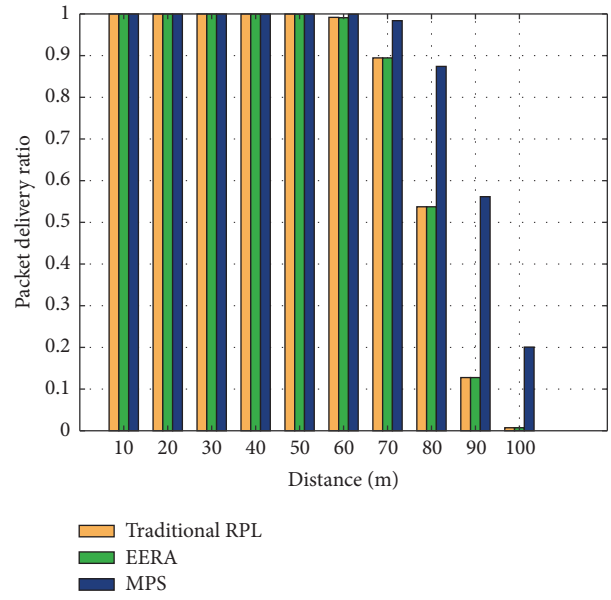


Figure 15: Average PDR of MPS as compared to RPL and EERA during energy usage data transmission.

Table 3: Percentage of network partitioning avoidance.

Rank	EERA	MPS
2	7.675%	5.65%
3	13.22%	16.27%
4	21.06%	25.63%
5	25.80%	32.97%
6	32.52%	39.32%
7	35.44%	44.30%
8	40.80%	48.00%
9	43.56%	52.47%

PDR as a function of distances. As a result, the size of packet affects the PDR. As compared to route construction period, the PDR during data transmission is lower than route construction at the same distance. As shown in Figure 15, by adopting CR technology, the PDR of MPS is much greater than traditional RPL and EERA. When MPS activates the CR mode, it transmits the data on the different spectrum frequency of Zigbee. Therefore, it has lower interference than using Zigbee channel. Moreover, based on different protocol, the transmission power on the CR mode is greater than Zigbee mode. Then, PDR can be improved.

Moreover, we also evaluate the ability of communication network partitioning avoidance as compared to the traditional RPL. As depicted in Table 3, by taking a residual energy into the parent selection account, MPS and EERA improve the performance of partitioned network avoiding of traditional RPL. Since MPS determines the weights using pair comparison between two performance metrics, MPS can avoid partitioned network better than EERA.

6. Conclusion

In this paper, a novel parent selection function—multicriteria parent selection (MPS) algorithm—is explored to address the issue of network partitioning and improves the transmission reliability of early parent selection functions. The three routing factors are used as the parent selection constraints; therefore, MPS can balance the workload of meters in AMI network and avoids the network partitioning efficiently. By considering the three routing factors as the parent selection constraints, the routing factors are analyzed more comprehensively than early parent selection algorithms. Moreover, we investigate the appropriate values for pairwise comparing between factors to generate the appropriate weights and those values are reported. Once the workload is balanced, the lifetime of meters in network is nearly the same and can be predicted easily. Then, it is fully facilitated to the maintenance team for planning the network maintenance time. Moreover, by adopting cognitive radio technology, the data is transmitted in an unused channel opportunistically where the channel quality is better than Zigbee. Therefore, MPS improves the transmission reliability and also reduces the number of data retransmissions.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] V. C. Gungor, D. Sahin, T. Kocak et al., "A survey on smart grid potential applications and communication requirements," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28–42, 2013.

[2] V. C. Güngör, D. Sahin, T. Kocak et al., "Smart grid technologies: communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.

[3] U. Ozgur, S. Tonyali, K. Akkaya, and F. Senel, "Comparative evaluation of smart grid AMI networks: performance under privacy," in *Proceedings of the 2016 IEEE Symposium on Computers and Communication, ISCC*, pp. 1134–1136, Italy, July 2016.

[4] L. Hernandez, C. Baladron, J. M. Aguiar et al., "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1460–1495, 2014.

[5] P. Sethi and S. R. Sarangi, "Internet of things: architectures, protocols, and applications," *Journal of Electrical and Computer Engineering*, vol. 2017, Article ID 9324035, pp. 1–25, 2017.

[6] D. F. Ramírez, S. Céspedes, C. Becerra, and C. Lazo, "Performance evaluation of future AMI applications in Smart Grid Neighborhood Area Networks," in *Proceedings of the 2015 IEEE Colombian Conference on Communications and Computing, COLCOM*, Colombia, May 2015.

[7] S. Nimbargi, S. Mhaisne, S. Nangare, and M. Sinha, "Review on AMI technology for Smart Meter," in *Communication and Computer Technology (ICAECCT*, pp. 21–27, Pune, 2016 IEEE International Conference on Advances in Electronics.

[8] D. Wang, Z. Tao, J. Zhang, and A. A. Abouzeid, "RPL based routing for advanced metering infrastructure in smart grid," in *Proceedings of the 2010 IEEE International Conference on Communications Workshops, ICC*, South Africa, May 2010.

[9] T. Winter, P. Thubert, A. Brandt et al., "RPL: IPv6 routing protocol for low-power and lossy networks," *J. Chem. Inf. Model*, vol. 53, no. 9, pp. 1689–1699, 2012, RFC6550, s.l.: IETF, http://tools.ietf.org/html/rfc6.

[10] I. Ishaq, D. Carels, G. Teklemariam et al., "IETF standardization in the field of the internet of things (IoT): a survey," *Journal of Sensor and Actuator Networks*, vol. 2, no. 2, pp. 235–287, 2013.

[11] A. Kheaksong, K. Srisomboon, and W. Lee, "A comparative study of various routing protocols for smart grid communication," *ECTI E-Magazine*, vol. 10, no. 1, 2016.

[12] H. Tian, Z. Qian, X. Wang, and X. Liang, "QoI-Aware DODAG Construction in RPL-Based Event Detection Wireless Sensor Networks," *Journal of Sensors*, vol. 2017, Article ID 1603713, 2017.

[13] P. Thubert, "Objective Function Zero for RPL," RFC 6552, 2012.

[14] A. S. Sadiq, T. Z. Almohammad, R. A. B. M. Khadri, A. A. Ahmed, and J. Lloret, "An Energy-Efficient Cross-Layer approach for cloud wireless green communications," in *Proceedings of the 2nd International Conference on Fog and Mobile Edge Computing, FMEC*, pp. 230–234, Spain, May 2017.

[15] W. Zhang, G. Han, Y. Feng, and J. Lloret, "IRPL: An energy efficient routing protocol for wireless sensor networks," *Journal of Systems Architecture*, vol. 75, pp. 35–49, 2017.

[16] A.-L. Kampen, K. Øvsthus, and Ø. Kure, "Energy balancing algorithms in wireless sensor networks," in *Proceedings of the Federated Conference on Computer Science and Information Systems, FedCSIS*, pp. 1223–1231, Poland, September 2015.

[17] O. Gnawali and P. Levis, "The Minimum Rank with Hysteresis Objective Function," Internet Requests for Comments RFC6719, 2012.

[18] E. Ancillotti, R. Bruno, and M. Conti, "Reliable data delivery with the IETF routing protocol for low-power and lossy networks," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 3, pp. 1864–1877, 2014.

[19] N. Pradeska, Widyawan, W. Najib, and S. S. Kusumawardani, "Performance analysis of objective function MRHOF and OF0 in routing protocol RPL IPV6 over low power wireless personal area networks (6LoWPAN)," in *Proceedings of the 8th International Conference on Information Technology and Electrical Engineering, ICITEE*, Indonesia, October 2016.

[20] L.-H. Chang, T.-H. Lee, S.-J. Chen, and C.-Y. Liao, "Energy-efficient oriented routing algorithm in wireless sensor networks," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, SMC*, pp. 3813–3818, October 2013.

[21] P. M. Esposito, M. E. M. Campista, I. M. Moraes, L. H. M. K. Costa, O. C. M. B. Duarte, and M. G. Rubinstein, "Implementing the expected transmission time metric for OLSR wireless mesh networks," in *Proceedings of the 1st IFIP Wireless Days (WD '08)*, pp. 1–5, IEEE, Dubai, UAE, November 2008.

[22] J. Mitola, "Cognitive radio architecture evolution," *Proceedings of the IEEE*, vol. 97, no. 4, pp. 626–641, 2009.

[23] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[24] S. Bayhan, S. Eryigit, F. Alagöz, and T. Tugcu, "Low complexity uplink schedulers for energy-efficient cognitive radio networks," *IEEE Wireless Communications Letters*, vol. 2, no. 3, pp. 363–366, 2013.

[25] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: An introduction to the first wireless standard based on cognitive radios," *Journal of Communications*, vol. 1, no. 1, pp. 38–47, 2006.

[26] N. R. Smith and K. Humood, *An efficient scheme in IEEE 802.22. WRAN for Real time and Non-Real time. Traffic delay*, Blekinge Institute of Technology, 2013.

[27] A. H. Process, "The Analytic Hierarchy Process," *Computing*, 1980.

[28] R. W. Saaty, "The analytic hierarchy process-what it is and how it is used," *Applied Mathematical Modelling: Simulation and Computation for Engineering and Environmental Systems*, vol. 9, no. 3-5, pp. 161–176, 1987.

[29] "Analytic hierarchy process (AHP) tutorial," [Online]. Available: http://people.revoledu.com/kardi/tutorial/AHP/AHP.htm.

[30] W. Lee, K. Srisomboon, and A. Prayote, "Fast spectrum sensing with coordinate system in cognitive radio networks," *ETRI Journal*, vol. 37, no. 3, pp. 491–501, 2015.

[31] P. Levis, T. Clausen, J. Hui, O. Gnawali, and J. Ko, "The Trickle Algorithm," Tech. Rep. RFC6206, 2011.

[32] S. Ajami and S. Ketabi, "Performance evaluation of medical records departments by analytical hierarchy process (AHP) approach in the selected hospitals in Isfahan: Medical Records Dep. & AHP," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1165–1171, 2012.

[33] Y. Min, L. T. Yang, F. Wang, and W. Wang, "Dynamic Sleeping Algorithm based on AHP for wireless sensor networks," in *Proceedings of the 2nd International Conference on Future Generation Communication and Networking, FGCN*, pp. 387–392, China, December 2008.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2976751, 9 pages
https://doi.org/10.1155/2018/2976751

Research Article

A Novel Indoor Positioning System Using Kernel Local Discriminant Analysis in Internet-of-Things

Sajida Imran and Young-Bae Ko

Department of Computer Engineering, Ajou University, Suwon 16499, Republic of Korea

Correspondence should be addressed to Young-Bae Ko; youngko@ajou.ac.kr

Received 8 September 2017; Revised 9 January 2018; Accepted 18 January 2018; Published 19 February 2018

Academic Editor: Waleed Ejaz

WLAN based localization is a key technique of location-based services (LBS) indoors. However, the indoor environment is complex; received signal strength (RSS) is highly uncertain, multimodal, and nonlinear. The traditional location estimation methods fail to provide fair estimation accuracy under the said environment. We proposed a novel indoor positioning system that considers the nonlinear discriminative feature extraction of RSS using kernel local Fisher discriminant analysis (KLFDA). KLFDA extracts location features in a well-preserved kernelized space. In the new kernel featured space, nonlinear RSS features are characterized effectively. Along with handling of nonlinearity, KLFDA also copes well with the multimodality in the RSS data. By performing KLFDA, the discriminating information contained in RSS is reorganized and maximally extracted. Prior to feature extraction, we performed outlier detection on RSS data to remove any anomalies present in the data. Experimental results show that the proposed approach obtains higher positioning accuracy by extracting maximal discriminate location features and discarding outlying information present in the RSS data.

1. Introduction

The outstanding advancement in IoT based applications has provoked the use of location-based systems (LBS) enabling mobile devices to provide a number of personal and commercial services including, but not limited to, object tracking [1], management and security, healthcare monitoring, personal navigation, and context awareness [2]. However, due to complex indoor environment with diverse requirements, despite research efforts for more than a decade, a widely deployed indoor localization system is not yet realized, which makes indoor localization an open research issue. Several candidate technologies are researched to solve indoor positioning problem including radio frequency (RF) [3], ultrawide band (UWB) [4], ultrasonic, and sound [5], visible light [6]. Most of these technologies provide comparatively accurate positioning. However, these technologies need specialized hardware to be deployed. Also, some of the technologies require line of sight (LOS) to work well, which is difficult to realize indoors. Accurate object localization is becoming more important for Wi-Fi based devices due to the increased use of augmented reality, social networking, health care monitoring, personal tracking, and other indoor location-aware applications.

The popularity and low price of Wi-Fi network interface cards are an attractive incentive to use Wi-Fi as the basis for a localization system. WLAN or Wi-Fi based positioning techniques locate the position of virtually every Wi-Fi compatible device without installing extra software or influencing the hardware. Wi-Fi has not been designed for positioning and yet the location can be estimated by leveraging RSS values on any WLAN equipped device without using any specialized hardware. The earnest advantage of Wi-Fi based indoor positioning over other indoor wireless technologies lies in its low cost and widespread deployment. The equipment is relatively cheap, while Wi-Fi infrastructure is the widely deployed communication infrastructure which saves time and money on new installations.

WLAN based positioning techniques are categorized into the following: Time of Arrival (ToA) or Time Difference of Arrival (TDoA) [7], Angle of Arrival (AoA) [8], and fingerprinting based technique [9]. ToA and AoA based

positioning techniques are difficult to implement indoors as these techniques require LOS measurement to work. Fingerprinting-based positioning on the other hand does not require any specialized hardware and makes use of widely deployed WLAN infrastructure. The fingerprinting localization system is based on the behavior of signal propagation and information about the geometry of the building to convert RSS values into distance values. This is a challenging task as, in an indoor environment, RSS values are affected by walls and obstacles which may reflect and propagate the signals, offering a nonlinear transformation between the RSS values and the physical location. Therefore, WLAN fingerprinting-based positioning should be modeled as a nonlinear and non-Gaussian dynamic system. In such case, nonlinear classifier and feature extraction method is preferred.

To improve positioning accuracy by minimizing the above described effects, a large number of access points (APs) are usually deployed. Increased number of APs can help in distinguishing more distinct locations. However, collecting RSS from all seen APs can create too many data dimensions and lead to the curse of dimensionality problem. In the curse of dimensionality, extracting useful information from high-dimensional data becomes restricted in case of limited training data which results in inaccurate position estimations. Remember that, due to changing indoor characteristics and labor intensity, the collected RSS training data are always limited. The curse of dimensionality can be handled by mapping high-dimensional data into low-dimensional data space to preserve inherent information. The solutions for dimensionality can be categorized into two classes: AP selection [10] and feature extraction [11]. In AP selection methods, the most important APs, that is, a subset of all APs, are selected. The AP selection method is less complicated by having RSS values of stable APs only. However, this method limits the classification performance with few distinct RSS values. In feature extraction and dimensionality reduction methods like Linear Discriminant Analysis (LDA) [12] and Principal Component Analysis (PCA) [13], RSS data from all seen APs are used to increase positioning accuracy. In PCA, RSS values are transformed into principal components. It searches for directions in the data that have largest variance and subsequently projects the data onto it. PCA improves performance by reducing the noise level. However, it mainly reduces the dimensionality of data without reducing variability in the data which makes it an inappropriate method for classification. LDA, on the other hand, provide more separate embedding for classification. LDA pick a new dimension that gives maximum separation between means of projected classes and minimum variance within each projected class. Fisher discriminant analysis (FDA), a variant of LDA, discriminates location features by maximizing between-class signal scatter and minimizes within-class signal scatter. However, feature extraction methods like PCA and FDA works only with unimodal and linear data and do not consider the multimodal and nonlinear RSS data. Local FDA (LFDA) [14] can handle multimodality by maximally preserving the local structure of the data. However, if the data is multimodal and nonlinear, direct mapping of signal data into physical location leads to very inaccurate position estimations. Kernel methods [15, 16] are used to extend the linear algorithms into equivalent nonlinear space. Therefore, we applied kernel method on LFDA to convert nonlinear space into kernel space where discriminative location features are extracted from the data.

In this work, we used KLFDA to cope with the multimodality and nonlinearity in RSS fingerprint data. KLFDA preserve the multimodal structure of the nonlinear data and provide more separate embedding than LFDA. The KLFDA's advantage over other kernel-based methods lies in its computational simplicity.

In addition to nonlinearity and multimodality, we also dealt with inordinate errors called outliers in RSS data. Due to interfering indoor environment, RSS data are prone to frequent outliers. Position estimation using these outlying RSS data leads to huge positional errors. Despite its huge impact on positioning accuracy, a little attention has been given to this issue. We used Maximum Likelihood Outlier Detection (MLOD) as an outlier detection algorithm, which is an inlier-based outlier detection algorithm.

In the proposed system, we first perform outlier detection to remove any anomalies present in RSS data. Then, for feature extraction, we mapped the raw RSS vectors into kernel feature space applying KLFDA (KLFDA transformation). Location is estimated by measuring Euclidean distance between kernelized offline and online RSS data features. Through experimental results, we witness the higher position accuracy of the proposed system.

The rest of the paper is organized as follows: Section 2 describes the background study on indoor positioning. The proposed kernel localized FDA is explained in Section 3. In Section 4 evaluation and results are provided. Finally, the conclusion is presented in Section 5.

2. Background Study

WLAN fingerprinting positioning consists of two phases: offline and online phase. In the offline phase, RSS values from multiple APs are collected at different reference points (RPs), that is, fingerprints to create the radio map. Furthermore, based on the collected fingerprints, a positioning model is trained to construct the relationship between RSS signals and physical locations. In the online phase, the user's location is estimated by applying the learned model to real-time RSS samples. Use of large RSS samples can help in creating more distinct location features. However, as a result the dimensionality of data is also increased, which leads to the misled feature classifications. Dimensionality reduction techniques like FDA work well by restricting the RSS data to certain low dimensions. However, the FDA only works with unimodal and linear data. The indoor environment is complex and, due to effects like multipath propagation, the RSS data become nonlinear and multimodal. When the FDA is applied on a multimodal and nonlinear data, it will form several different clusters from a single multimodal sample. It is difficult for FDA to keep within-class scatter to a certain level. For dimensionality reduction of multimodal data, the local structure of data needs to be preserved. LFDA does not require multimodal samples to fall into a single cluster.

As a result, more degree of freedom is left for increasing separability. LFDA can preserve multimodal structure of the data better than FDA. However, LFDA also works on linear data structures only.

To deal with multimodality in the data, Sugiyama [14] proposed a dimensionality reduction method called enhanced local Fisher discriminant analysis (ELFDA) which is a localized version of FDA. It takes local structure of the data into account so that the multimodal data can be embedded appropriately. ELFDA can preserve multimodal structure of the data better than FDA and provides more separate embedding than FDA. However, ELFDA only works in linear data structures. Deng and Meng in [16] used kernel Fisher discriminant analysis (KFDA) for indoor localization. KFDA is meant to be use for classifying offline RSS measurements with online RSS measurements. However, the work does not consider any localized variant of FDA to consider multimodal data. In [17], LFDA based indoor positioning is achieved. The work considers the multimodality in data and used LFDA to cope with the problem. A cluster-based approach is used to minimize the search space to a specific cluster. However, this approach did not consider any nonlinearity present in the RSS data. The study [17] used LFDA to cope with multimodality in the data. LFDA is used to extract discriminative features from the RSS data. These discriminative features are extracted in such a way to increase between-class separability, while preserving within-class local structure of the RSS space. The generalization ability of LFDA is further enhanced using signal perturbation, which generates a higher number of representative training samples.

In Hayashi et al. [18], a fingerprinting-based Wi-Fi indoor positioning method robust against temporal fluctuations and spatial instability in Wi-Fi signals is presented. Spatial changes are coped with by splitting the environment of interest into several areas and tailoring several weak estimators to each area so that they can accurately estimate the user's position in the given area. They cope with temporal changes in the Wi-Fi signals by using random subsets of APs in weak estimators. Zheng et al. in [19] consider two problems commonly found in RSS data: unstable positioning fingerprint features and curse of dimensionality. They designed a positioning fingerprint feature using the segment similarity of Wi-Fi access points by considering both the received signal strength value and the Wi-Fi access point. Based on this designed fingerprint feature, a two-stage positioning algorithm for indoor fingerprint-based positioning is proposed. Pan et al. in [20] considered uncertainty and nonlinearity in the signal to map signal and physical space using multidimensional vector regression and kernel method. For feature extraction, kernel canonical analysis is performed to maximally correlate pairwise similarity in both signal and physical space.

In Khalamehrabadi et al. [21], a WLAN based localization scheme is proposed that consists of outlier detection and radio-map interpolation schemes. Outlier detection is mapped as augmented optimization problem. A GS-based positioning system is reformulated to cope with outliers during online phase only. The work [22] proposed a joint WLAN based localization and outlier detection scheme. The scheme consists of three phases of course localization, AP selection, and fine localization using sparse recovery algorithms. The working area is clustered into ROIs where user location is searched within a ROI. Outlier detection is performed during online phase using modified sparse recovery algorithm.

Chen and Juang in [23] proposed an outlier detection framework to cope with the outliers in data for localization. The scheme does nothing to remove the outliers and rather repairs the outliers to preserve useful information contained with outliers. Each piece of outlier data is given a confidence value based on MAD-scale score which is introduced to show the usefulness of the data. The proposed outlier detection based scheme is applied in RSS data collection and location tracking phases. Meng et al. in [24] developed a secure and robust indoor localization scheme to address the outliers in RSS data due to accidental environmental changes and access point attacks. A probabilistic region-based fingerprinting method is proposed to reduce the outlier effect and improve the localization accuracy. A three-step location sensing algorithm is proposed.

Proposed positioning system works on both multimodality and nonlinearity in RSS data. Multimodality in data is handled with localized FDA, while kernel method is used to deal with nonlinearity in RSS fingerprinting data. Furthermore, MLOD based outlier detection is performed to remove any anomalies (also called noise) in RSS data. We experienced that the use of outlier detection algorithm prior to feature extraction facilitates more accurate classification of distinguished location features. Experimental results show improved position estimation accuracy as compared to other feature extraction methods of LDA [12], PCA [13], and ELFDA [14].

3. Proposed KLFDA Positioning Algorithm

This section describes the proposed KLFDA based positioning algorithm. We first explain the RSS data collection and radio-map generation step. Then, overview of KLFDA is presented along with feature extraction in nonlinear data. Afterwards, distance calculation of RSS vector in new feature space is described to estimate the user's final position. Choice of Kernel selection is explained afterwards. Finally, outlier detection using MLOD is presented that removes any anomalies present in RSS data to make the KLFDA based positioning more efficient.

3.1. Overview of KLFDA. In offline phase of the proposed KLFDA, the following sequence of steps is followed: at first, a road map is built by collecting RSS samples at different reference points. Then, MLOD is performed to remove any anomalies present in the data. After that, for feature extraction, we first mapped the raw RSS vectors into kernel feature space by applying KLFDA. This process is referred to as KLFDA transformation.

In online phase of the proposed KLFDA, the following sequence of steps is followed: after collection of real-time RSS samples, the outlier detection and KLFDA transformation are applied in the same way as applied in offline data. Then, location is estimated by measuring Euclidean distance

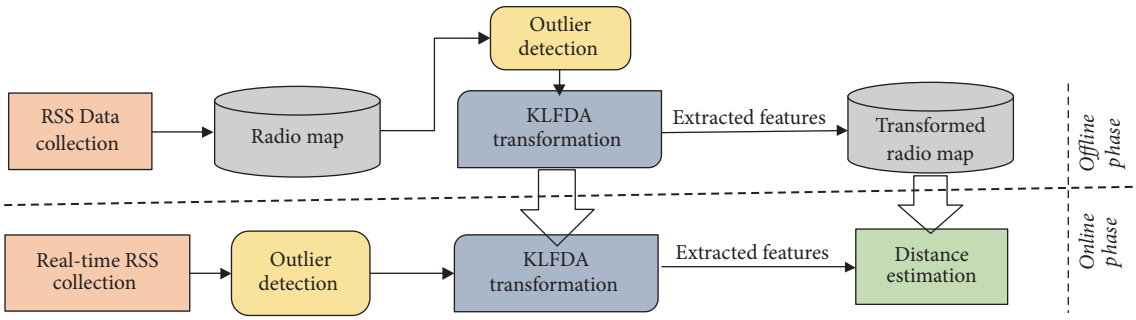


Figure 1: Proposed framework.

between kernelized offline and online RSS data features. Figure 1 shows the proposed indoor positioning framework that tackles multimodal and nonlinear RSS data.

RSS data is multimodal, non-Gaussian, and nonlinear in nature. Kernel methods implicitly map linear data into high-dimensional nonlinear feature space. KLFDA can well preserve the multimodal structure of the nonlinear data and provide more separate embedding than its nonkernel variants.

3.2. Feature Generation. FDA finds a linear combination of features that characterizes or separates two or more classes of objects or events. We used localized FDA (LFDA) to look after the multimodality and then applied kernel method on LFDA to cope with nonlinearity in RSS data (KLFDA). LFDA maximally preserve local structure of the RSS class data than simple FDA which result in better separation on multimodal data. The resulting combination is used as a linear classifier or more commonly for dimensionality reduction. To extract location feature to reduce RSS data dimensionality, the data is modeled using FDA. Let $a_i \in \mathbb{R}^m$ $(i = 1, 2, 3, \ldots, n)$ be m-dimensional RSS samples and $b_i \in \{1, 2, 3, \ldots, c\}$ are related class labels, where n is number of samples and c is the number of classes. Let n_s be the number of samples in class s:

$$\sum_{s=1}^{c} n_s = n. \tag{1}$$

Let X be the matrix of all samples:

$$X \equiv (x_1 \mid x_2 \mid \cdots \mid x_n). \tag{2}$$

Let $z_i \in \mathbb{R}^r$ $(1 \leq r \leq m)$ be the low-dimensional representations of a_i, where r is the reduced dimension. Persuasively, we consider m to be large and r to be small, but not limited to such cases. For linear dimensionality reduction, the embedded samples z_i are given as

$$z_i = T^{\top} x_i, \tag{3}$$

where $\top$ denotes the transpose of the RSS matrix. The goal of LFDA is to find a linear transformation that maximizes class separability in the reduced dimensional space. Let us have $S_{(b)}$ as between-class scatter and $S_{(w)}$ as within-class scatter. To formulate RSS database, the RSS vector is represented as

$$A = [a_1, a_2, \ldots, a_n], \tag{4}$$

where a_1, a_2 represent the received signal strength at point P from AP_i. The RSS vector $A = [a_1, a_2, \ldots, a_n]$ is formulated with FDA in such a way to maximize

$$\begin{aligned} j(u) &= \frac{u^{\top} S_{(b)} u}{u^{\top} S_{(w)} u}, \\ S_{(w)} &= \frac{1}{A} \sum_{l=1}^{c} \sum_{i:y_{i=l}} (x_i - u_l)(x_i - u_l)^{\top}, \\ S_{(b)} &= \frac{1}{A} \sum_{l=1}^{c} n_l (u_l - u)(u_l - u)^{\top}, \end{aligned} \tag{5}$$

where $x_i = (1/n_i) \sum_{j \in N_i} a_j$ is class centroid while $u = (1/n) \sum_{j=1}^{n} a_j$ is global centroid. $\sum_{i:y_{i=l}} \cdot$, which represents the summation over i, that is, $y_i = l, u_l$ is the mean of the samples in class l, and u is the mean of all samples. FDA's transformation matrix T maximizes the between-class scatter, while it minimizes the within-class scatter, where FDA transformation T is defined as follows:

$$T \equiv \underset{T \in R^{d \times r}}{\arg} \left[\left(T^{\top} S_{(b)} T \left(T^{\top} S_{(w)} \right)^{-1} \right)^{\top} \right]. \tag{6}$$

While performing classification on multimodal data, FDA restricts within-class scatter to be small which may result in combining multimodal data into a single cluster. Smaller within-class distance can restrict the increase in feature separability which could result in degraded classification ability of the FDA. Local FDA, on the other hand, works locally and does not necessarily impose the restriction of RSS samples to be close, which gives an increase in between-class distance, resulting in the extraction of more discriminating location features. List of symbols used in this article is listed in the Symbols.

3.2.1. Kernel Local Fisher Discriminant Analysis. LFDA works well to cope with multimodal linear RSS data. However, RSS data usually show nonlinear behavior, especially at points of sudden turns. Kernel methods are proven to deal well with the nonlinearity in the data. To extend LFDA to kernelized LFDA for nonlinear mapping, RSS data can be mapped to

new kernel feature space $\dot{F}$ using function ϕ. We want to maximize the function

$$j(u) = \frac{u^\top S^{\phi}_{(b)} u}{u^\top S^{\phi}_{(w)} u}, \tag{7}$$

in which now $u \in F$ and $S^{\phi}_{(b)}$ and $S^{\phi}_{(w)}$ are the respected matrices in $\dot{F}$ which is represented as

$$S^{\phi}_{(w)} = \frac{1}{A}\sum_{l=1}^{c}\sum_{i:y_{i=l}} (x_i - u_l)(x_i - u_l)^\top,$$
$$S^{\phi}_{(b)} = \frac{1}{A}\sum_{l=1}^{c} n_l (u_l - u)(u_l - u)^\top. \tag{8}$$

The optimal discriminant basis vectors are obtained by solving eigenvalue problem $u^* = \arg\max(j(u))$ in (7). The kernel feature representation of corresponding RSS samples can be collected by the projection of kernel representation onto the optimal discriminant basis vectors u. The optimal value of u that maximizes $j(u)$ can be obtained by the intersection space of the null space of $S^{\phi}_{(w)}$ and the nonzero space of $S^{\phi}_{(b)}$, where nonzero space of $S^{\phi}_{(b)}$ can be procured by keeping the eigenvectors $M = [m_1 \cdots m_E]$ with the E biggest eigenvalues $[\lambda_1 \cdots \lambda_E]$. Smaller eigenvalue vectors are considered as noise as these vectors contain little discriminative information.

3.3. Choice of Kernel. In indoor positioning, kernel methods are popular to implicitly map linear data into high-dimensional nonlinear feature space. For a given kernel function, the corresponding underlying nonlinear transformation function is determined which defines the mapping from original data space to kernel induced feature space. Therefore, it can be said that the form of kernel function plays a very important role in kernel methods. In this work, the choice of kernel mainly relies on the nonlinearity and uncertainty properties of RSS. Kernel parameters are adjusted through a process called generalized-cross-validation. We are interested in kernel K for RSS examples a and b in an input feature space $X = \mathbb{R}^d$ as

$$K(a, b) = \langle \emptyset(a), \emptyset(b) \rangle, \tag{9}$$

where $\emptyset$ nonlinearly maps linear input space X into linear feature space F. Generally, kernels are categorized into stationary, locally stationary, and nonstationary kernels. Stationary kernels, sometimes called anisotropic stationary kernel, depend on lag vector that separates RSS examples a and b. At a fixed location, RSS from one AP may vary as high as 10 dB and always shows Gaussian or semi-Gaussian distribution. Among many, Gaussian kernels are popular in RSS based localization because of their good localization and smoothing ability. Moreover, Gaussian kernels are good at characterizing the uncertainty of RSS and ease in capturing the nonlinear RSS patterns in the Gaussian kernel induced space. Therefore, we chose Gaussian kernel for nonlinear mapping, and a 2D Gaussian kernel is defined as

$$G(a, b, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{a^2 + b^2}{2\pi\sigma^2}\right), \tag{10}$$

where a and b are two RSS vectors and σ determines the kernel width. σ is a similarity measure between two RSS vector samples. Also, it greatly affects shape of the kernelized feature space by controlling the sensitivity of the kernel to change. In this work we used method from [15] to determine kernel width.

3.4. Outlier Detection. Due to variability in the measurements of a dataset, there could be observations that are different from other observations present in the data. These different observations are called outliers. Outlier in a dataset are present due to a number of reasons like shadowing and multipath and variability in transmitting power of APs to manage network traffic; if traffic loads during offline and online data collection are different, then the RSS readings will certainly be different. Also, due to impermanent effects in any AP, the RSS readings may not be available in any of the offline or online phases. The described phenomena generate outliers in the RSS data which results in difference in offline and online RSS measurements. This difference can significantly influence position estimation accuracy. In this paper we adopt Maximum Likelihood Outlier Detection, which is an inlier-based outlier detection algorithm that works on detecting outliers in sample data based on some model dataset. The sample dataset (evaluation set) is organized based on degree of outlining. The degree of outlyingness is measured as

$$f = \frac{E_s}{M_s}. \tag{11}$$

The ratio is estimated by the density-ratio estimation method KL importance estimation procedure (KLIEP). KLIEP's performance is estimated through basic function $\Psi(x)$ [19]. In such case, use of Gaussian kernel is preferred which is expressed as follows:

$$r(x) = \sum_{l=1}^{n_{nu}} \emptyset_l K\left(x, {x_l}^{nu}\right), \tag{12}$$

where $K(x, x_l)$ represents Gaussian kernel. Choosing large n_{nu} needs Gaussian centers to be chosen by using all $\{x_i^{nu}\}_{i=1}^{n_{nu}}$ samples, which is computationally expensive. This problem can be solved by using subset of $\{x_i^{nu}\}_{i=1}^{n_{nu}}$ as Gaussian centers, which is expressed as follows:

$$r(x) = \sum_{l=1}^{z} \emptyset_l K(x, c_l), \tag{13}$$

where c_l is a randomly chosen template point from $\{x_i^{nu}\}_{i=1}^{n_{nu}}$ and z ($\in \{1, \ldots, n_{nu}\}$) represents a prefix number. KLFDA transforms raw RSS data into high-dimensional kernel space, which contains enhanced nonlinear discriminative location information. Moreover, it also adheres to Fisher criterion, which maximizes the between-class scatter while it minimizes within-class scatter at the same time. Distance calculation between offline and online RSS data is then applied for location estimation. It is achieved by matching the k-nearest RSS vectors through KLFDA distance measure.

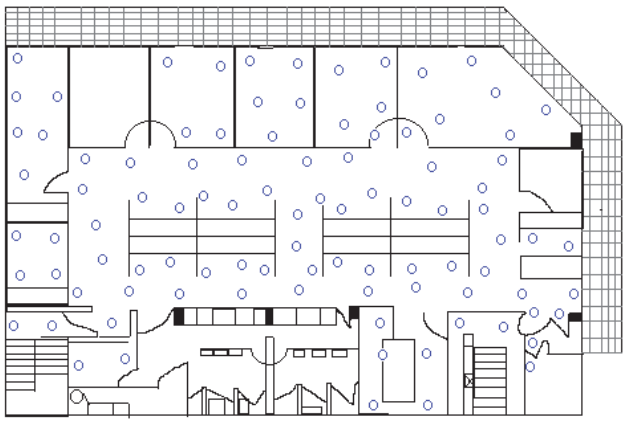


Figure 2: Experimental setup. Purple circles show the reference locations used during offline phase.

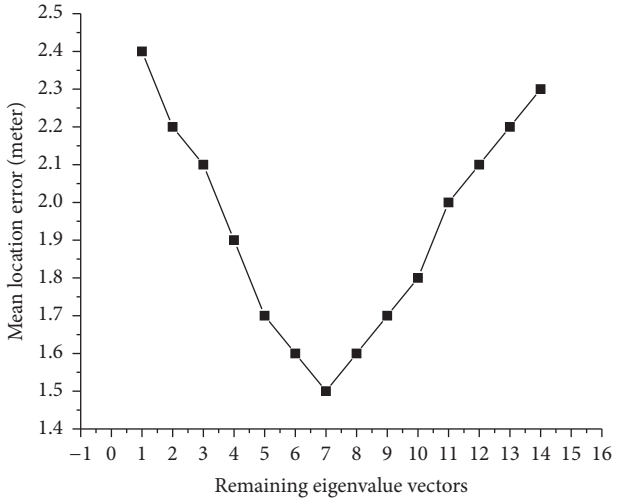


Figure 3: Effect of different number of remaining eigenvectors E on the mean location error.

4. Evaluation

4.1. Experimental Setup. We used KIOS RSS dataset to train and validate our positioning system. Wi-Fi RSS data is collected through extensive measurements at KIOS Research Center, Cyprus [25]. The area consists of a 560 m^2 typical office environment that includes a conference room, several open cubicles and private offices, laboratories, and corridors. Total 9 stable APs with full coverage were found to collect RSS data at the floor. Moreover, varying number of unstable APs were found at different points in corridors during RSS data collection. Five different devices were used to collect RSS data at different points. The mobile devices used for RSS data collection are an Asus PC T101MT laptop running Windows 9, a HP iPAQ hw6915 PDA with Windows Mobile, an HTC Flyer Android tablet, HTC Desire, and Samsung Nexus S. The area is divided into 201 reference points. Out of 201 reference points, 2100 RSS readings at 105 reference points are used to train the model with 20 RSS readings per reference point, while 960 RSS readings at 96 with 10 readings per reference points are used to test the model. Figure 2 shows the schematics of the experimental setup. The proposed approach is compared with PCA [12], LDA [11], and ELFDA [13]. The whole training set for radio map is divided into five parts: four parts are for tentatively building the model, while the remaining one is a validation set for evaluating the positioning performance. Each part is chosen as a validation set for one time and the validation performance is the average result of five times. The real model is set by parameter values with the found best validation performance. The parameters of the other compared algorithms are all set to the found optimal values by this method.

4.2. Analysis of Simulation Results. Maximally extracted location features can better classify individual locations. For accurate position estimation, it is important to fully extract discriminative features in RSS data. This discriminating quantity is obtained by the cumulative percentage of extracting eigenvalues compared with the remaining eigenvectors E. Figure 3 compares accuracy of different feature generation

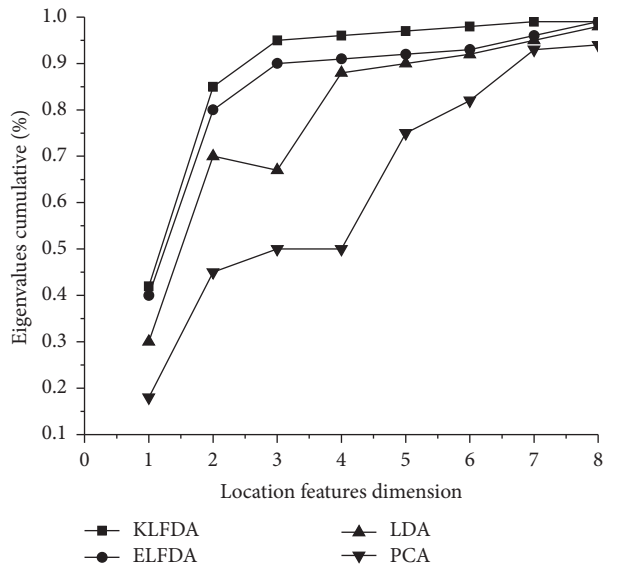


Figure 4: Cumulative eigenvalues percentage against feature dimensions.

methods. It is shown in the figure that the mean location error is the lowest when E is equal to 7, which shows that 97.0% nonlinear information is retained. The remaining 3.0% is considered as noise which is discarded. We see a decrease in mean location error until within-class scatter reaches 7, and after that the mean error starts increasing. The eigenvectors with larger eigenvalues are discarded since these larger values increase the within-class scatter which can result in degraded positioning accuracy. The proposed approach can maximally extract nonlinear distinct location feature. The information that increases within-class scatter or decreases between-class scatter is considered noise. Although extremely larger between-class scatter can lead to increased uncertainty, however, marginally higher between-class scatter preserves more location information. Figure 4 shows the eigenvalues cumulative percentage against feature

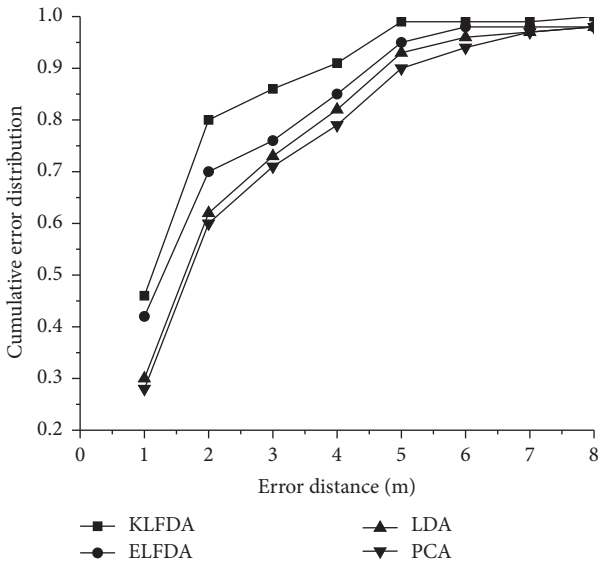


Figure 5: Accuracy comparison between different feature generation methods.

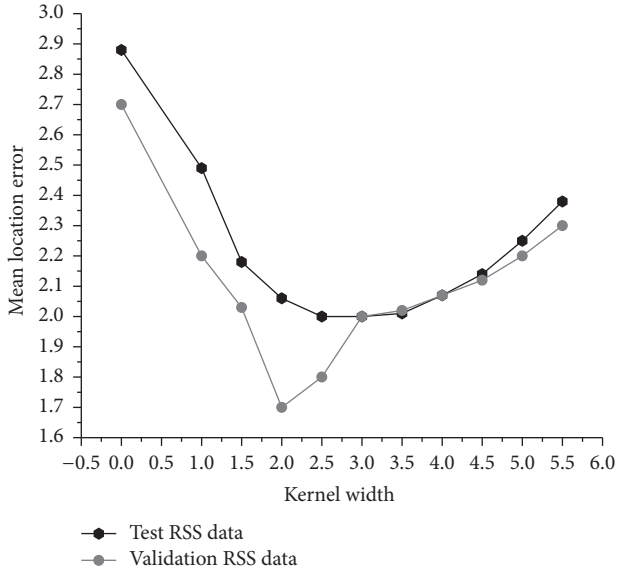


Figure 6: Kernel width versus mean location error.

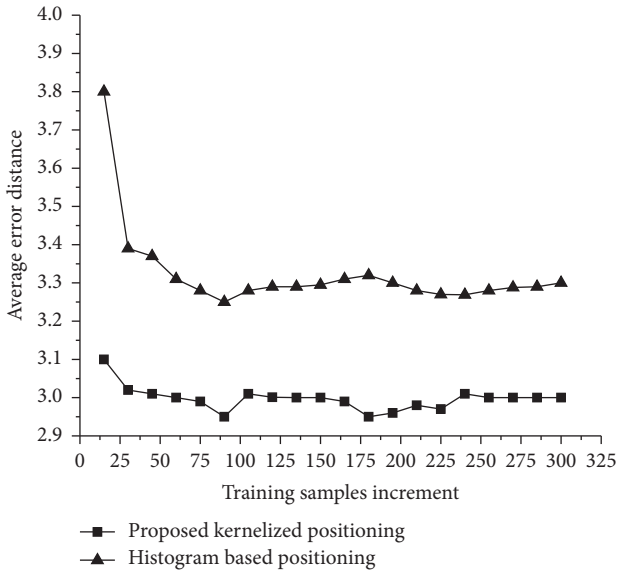


Figure 7: Number of training samples versus average error distance.

dimensions. The proposed approach is compared with ELFDA, LDA, and PCA. PCA can maximize the variance of extracted features, whereas ELFDA and LDA work on maximizing the between-class separability. However, we see in the figure that the proposed KLFDA preserves local within-class structure without being effected by nonlinearity in RSS data. The first three dimensions of KLFDA already have a cumulative percentage of more than 90% as compared with ELFDA, LDA, and PCA which indicates its strongest discriminative power. Figure 5 shows the average testing positioning accuracy that indicates the cumulative positioning error distribution. Accuracy is also compared between PCA, LDA, ELFDA, and proposed KLFDA. Error distance is the Euclidean distance between true and estimated location coordinates. The cumulative probability distribution of error distance is used as positioning accuracy. From the figure we see that, within error distance of 2 m, the accuracy of the proposed KLFDA is 80%, while accuracy of ELFDA, LDA, and PCA is 70%, 62%, and 60%, respectively.

The shape of the kernelized feature space is mainly defined by σ which is a parameter to restraint the kernel width. Determining the parameter σ is a complex task. Figure 6 shows the mean location error for different kernel widths because of constraints in computations and facing the variances between offline and online measurements. Near-optimal kernel width is observed in validation data for the test data, which encourages urging for parameter tuning on valid validation data. At σ value 2, the larger gap between validation and test data is because of larger variance between offline and online measurements.

While computing distance, kernel-based system's computation complexity is increased. Smaller test samples result in less calibration time in offline phase. Figure 7 shows the effect of number of training samples on distance error. Comparison is made between well-known histogram positioning method [26] and the proposed scheme. Histogram is a popular technique of estimating PDF in a nonparametric way. The figure shows the number of samples collected at each reference point. We see that histogram method shows larger variance with number of samples. This is because a substantial number of samples are required at each training location to produce satisfactory results in histogram based methods. The proposed approach on the other hand shows almost consistent results with the increased number of samples, which shows the invariability of the proposed scheme against number of samples. We also see that both the systems show highest accuracy with almost 90 training samples. Figure 8 shows the mean location error observed with increasing number of outliers. It is shown in the figure that KLFDA with outlier detection has mean location error less than that of KLFDA where there is no outlier detection. The results

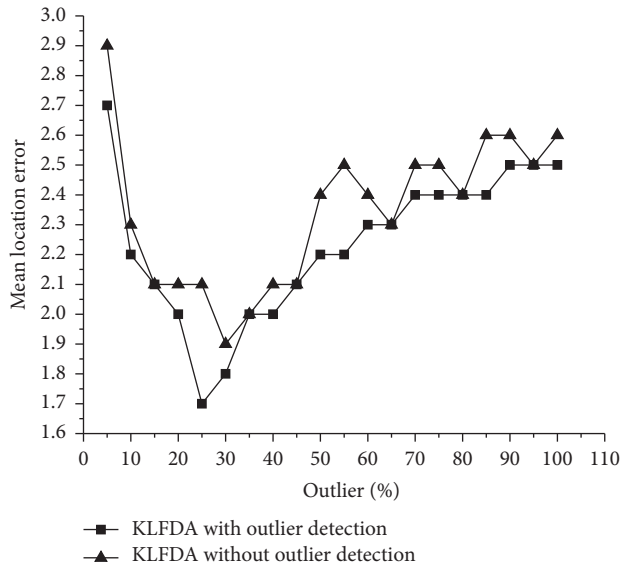


Figure 8: Increase in outliers' versus mean location error.

of Figure 8 acknowledge usefulness of outlier detection in proposed scheme.

5. Conclusion and Future Work

Accuracy of fingerprinting-based localization increases with higher number of collected WLAN signals. Due to complex environment, the collected RSS values are highly uncertain, multimodal, and nonlinear in nature, which makes the positioning system perform inaccurately. We performed nonlinear discriminative feature extraction of RSS using KLFDA which extracts location features in a kernel space where the nonlinear RSS features are well characterized and captured. Along with handling nonlinearity in data, KLFDA copes well with the multimodality in the RSS data. By performing KLFDA, the discriminative information contained in RSS is reorganized and maximally extracted. Outlier detection on RSS data removes anomalies in the data which makes KLFDA based location system efficiently perform position estimation. The proposed approach obtains higher accuracy by maximally extracting discriminative features in nonlinear space. We aim to extend this work by performing extensive experiments to thoroughly analyze further discriminative features and effects of outlier detection on feature extraction efficiency.

Symbols

a_i: m-dimensional RSS samples
b_i: Class labels
n: Number of samples
c: Number of classes
n_s: Number of samples in class s
X: Matrix of all samples
z_i: Low-dimensional representation of a_i
r: Reduced dimension
$S_{(b)}$: Between-class scatter
$S^{\Phi}_{(b)}$: Between-class scatter in kernelized feature space
$S_{(w)}$: Within-class scatter
$S^{\Phi}_{(w)}$: Within-class scatter in kernelized feature space
T: FDA transformation matrix
$\dot{F}$: Kernel feature space
ϕ: Function to map RSS vectors to new kernel feature space
E_s: Sample/evaluation dataset
M_s: Model dataset.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01059049).

References

[1] S. Imran and Y.-B. Ko, "A continuous object boundary detection and tracking scheme for failure-prone sensor networks," *Sensors*, vol. 17, no. 2, 2017.

[2] J. Chon and H. Cha, "LifeMap: a smartphone-based context provider for location-based services," *IEEE Pervasive Computing*, vol. 10, no. 2, pp. 58–67, 2011.

[3] S. He and S. G. Chan, "Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[4] S. Gezici, Z. Tian, G. B. Giannakis et al., "Localization via ultra-wideband radios: a look at positioning aspects of future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, 2005.

[5] F. Ijaz, H. K. Yang, A. W. Ahmad, and C. Lee, "Indoor positioning: a review of indoor ultrasonic positioning systems," in *Proceedings of the 15th International Conference on Advanced Communication Technology: Smart Services with Internet of Things!, ICACT 2013*, pp. 1146–1150, 2013.

[6] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang, "Wearables can afford: light-weight indoor positioning with visible light," in *Proceedings of the the 13th Annual International Conference*, pp. 317–330, Florence, Italy, 2015.

[7] Y. Qi, H. Kobayashi, and H. Suda, "Analysis of wireless geolocation in a non-line-of-sight environment," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 672–681, 2006.

[8] C. Wong, R. Klukas, and G. Messier, "Using WLAN infrastructure for angle-of-arrival indoor user location," in *Proceedings of the 68th Semi-Annual IEEE Vehicular Technology (VTC '08)*, pp. 1–5, IEEE, 2008.

[9] S. Chan and G. Sohn, "Indoor localization using wi-fi based fingerprinting and trilateration techiques for LBS applications," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-4/C26, pp. 1–5, 2012.

[10] H. Miao, Z. Wang, J. Wang, L. Zhang, and L. Zhengfeng, "A novel access point selection strategy for indoor location with Wi-Fi," in *Proceedings of the 26th Chinese Control and Decision Conference, CCDC 2014*, pp. 5260–5265, Changsha, China, 2014.

[11] Z.-A. Deng, Y.-B. Xu, and L. Ma, "Indoor positioning via nonlinear discriminative feature extraction in wireless local area network," *Computer Communications*, vol. 35, no. 6, pp. 738–747, 2012.

[12] C. H. Park and H. Park, "A comparison of generalized linear discriminant analysis algorithms," *Pattern Recognition*, vol. 41, no. 3, pp. 1083–1097, 2008.

[13] S.-H. Fang and T. Lin, "Principal component localization in indoor wlan environments," *IEEE Transactions on Mobile Computing*, vol. 11, no. 1, pp. 100–110, 2012.

[14] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 905–912, ACM, 2006.

[15] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689–705, 2007.

[16] Y. Xu, Z. Deng, and W. Meng, "An indoor positioning algorithm with kernel direct discriminant analysis," in *Proceedings of the 53rd IEEE Global Communications Conference (GLOBECOM '10)*, pp. 1–5, Miami, Fla, USA, December 2010.

[17] Z.-A. Deng, Y. Xu, and L. Chen, "Localized local fisher discriminant analysis for indoor positioning in wireless local area network," in *Proceedings of the 2013 IEEE Wireless Communications and Networking Conference, WCNC 2013*, pp. 4795–4799, Shanghai, China, 2013.

[18] T. Hayashi, D. Taniuchi, J. Korpela, and T. Maekawa, "Spatio-temporal adaptive indoor positioning using an ensemble approach," *Pervasive and Mobile Computing*, vol. 41, pp. 319–332, 2017.

[19] Z. Zheng, Y. Chen, T. He, L. Sun, and D. Chen, "Feature learning for fingerprint-based positioning in indoor environment," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 452590, 2015.

[20] J. J. Pan, J. T. Kwok, Q. Yang, and Y. Chen, "Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 9, pp. 1181–1193, 2006.

[21] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Structured group sparsity: a novel indoor wlan localization, outlier detection, and radio map interpolation scheme," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6498–6510, 2017.

[22] A. Khalajmehrabadi, N. Gatsis, D. J. Pack, and D. Akopian, "A Joint Indoor WLAN Localization and Outlier Detection Scheme Using LASSO and Elastic-Net Optimization Techniques," *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2079–2092, 2017.

[23] Y.-C. Chen and J.-C. Juang, "Outlier-detection-based indoor localization system for wireless sensor networks," *International Journal of Navigation and Observation*, vol. 2012, Article ID 961785, 2012.

[24] W. Meng, W. Xiao, W. Ni, and L. Xie, "Secure and robust Wi-Fi fingerprinting indoor localization," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '11)*, pp. 1–7, IEEE, Guimarães, Portugal, 2011.

[25] C. G. P. Christos Laoudias and R. Piché, *KIOS WiFi RSS Dataset*, https://www.researchgate.net/publication/256482916_KIOS_WiFi_RSS_dataset.

[26] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6839671, 13 pages
https://doi.org/10.1155/2018/6839671

Research Article

EE-MRP: Energy-Efficient Multistage Routing Protocol for Wireless Sensor Networks

Muhammad Kamran Khan,[1] Muhammad Shiraz,[1] Kayhan Zrar Ghafoor,[2,3] Suleman Khan,[4] Ali Safaa Sadiq,[4] and Ghufran Ahmed[5]

[1]*Department of Computer Science, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan*
[2]*Department of Computer Science, Faculty of Science, Cihan University-Erbil, Kurdistan Region, Iraq*
[3]*Department of Software Engineering, College of Engineering, Salahaddin University-Hawler, Kurdistan Region, Iraq*
[4]*School of Information Technology, Monash University, 47500 Bandar Sunway, Malaysia*
[5]*Department of Computer Science, COMSATS Institute of Information Technology, Islamabad, Pakistan*

Correspondence should be addressed to Suleman Khan; suleman.khan@monash.edu

Received 11 August 2017; Revised 27 October 2017; Accepted 19 November 2017; Published 8 January 2018

Academic Editor: Syed Hassan Ahmed

Wireless sensor networks (WSNs) have captivated substantial attention from both industrial and academic research in the last few years. The major factor behind the research efforts in that field is their vast range of applications which include surveillance systems, military operations, health care, environment event monitoring, and human safety. However, sensor nodes are low potential and energy constrained devices; therefore, energy-efficient routing protocol is the foremost concern. In this paper, an energy-efficient routing protocol for wireless sensor networks is proposed. Our protocol consists of a routing algorithm for the transmission of data, cluster head selection algorithm, and a scheme for the formation of clusters. On the basis of energy analysis of the existing routing protocols, a multistage data transmission mechanism is proposed. An efficient cluster head selection algorithm is adopted and unnecessary frequency of reclustering is exterminated. Static clustering is used for efficient selection of cluster heads. The performance and energy efficiency of our proposed routing protocol are assessed by the comparison of the existing routing protocols on a simulation platform. On the basis of simulation results, it is observed that our proposed routing protocol (EE-MRP) has performed well in terms of overall network lifetime, throughput, and energy efficiency.

1. Introduction

Wireless sensor network is comprised of a large number of tiny and small sensor nodes which are distributed over the physical environment to monitor security events, temperature, humidity, capture images, pressure, and so on [1–8]. Sensor nodes have limited energy capabilities and have individual resources (such as CPU and memory). These nodes are randomly located in the dynamically varying environment [9]. The life of the sensor node depends on the energy (battery) of the node, on which the lifespan of the network is dependent. The main problem faced in WSN is that the energy of sensor nodes dropped quickly and become lifeless [10]. It is observed that maximum energy is dissipated in the communication subsystem [11]. Therefore, in order to extend and maximize the lifetime of sensor nodes, designing energy-efficient algorithms is necessary [1, 11–13]. In order to increase the lifetime of WSN, it is needed to manage resources carefully.

There are many issues in WSNs which have to be considered, such as overall lifetime, coverage, energy efficiency, and network security [14–17]. Apart from efficient and reliable communication, the major goal of a routing protocol is to maximize the lifetime of cluster heads and sensor nodes. The following issues should be considered while designing an energy-efficient clustering algorithm. Firstly, the routing protocol should be distributed because it works better for large-scale WSNs [18]. Secondly, cluster heads should be distributed evenly in the network field, so that all sensor nodes equally find cluster heads for the communication

[19]. Thirdly, communication between cluster heads and base station should be minimized because maximum energy is utilized in the communication between cluster heads and base station [20]. Fourthly, in most of the hierarchical routing protocols, cluster heads selection technique is not efficient [21].

The energy consumption in the gathering of information from sensor nodes varies among cluster heads because it depends on the number of members of cluster heads. The consumption of energy also differs in cluster members because it depends on the distance between member nodes and cluster heads. In most of the existing WSN routing protocols, it is noticed that if one issue is addressed then other issues are ignored, due to which the required energy efficiency is not achieved. All factors discussed above should be considered in the development of routing protocol to achieve maximum energy efficiency in combination with the target to achieve maximum network coverage and data throughput.

The primary objective of this research is to address routing issues by adopting an efficient cluster head selection method and proposing an energy-efficient and reliable routing protocol for WSNs. A lot of cluster-based routing protocols have been proposed for WSNs, but these protocols have limitations due to challenges related to the determination of accurate radio model (communication model) for the sensor nodes and cluster heads in the network area. The other drawback which is also addressed in our research work is an uneven distribution of cluster heads in the network field, which results in disconnection of a portion of network area from the base station. Hierarchical routing protocols like LEACH [4] and its variants use the same amplification energy in the transmission of data from the source node to the destination regardless of the distance between transmitting and receiving nodes [4, 10, 22]. Energy can be preserved by using multiple energy levels for the transmission of data according to the distance between the receiver and the transmitter. Therefore, we need to propose an energy-efficient solution to maximize the overall lifetime of the network.

The rest of the paper is organized as follows. Section 2 presents related work. The proposed routing protocol is presented in Section 3. Section 4 presents performance evaluation in terms of energy efficiency, stability period, and throughput. Finally, Section 5 ends up with the conclusion.

2. Related Work

WSNs have massive complexity and applicability, because the complex nature of WSNs variety of issues has to be addressed by the scientists and engineers. Various energy-efficient routing protocols have been proposed in the last few years. In hierarchical cluster-based network, the network is divided into separate clusters, and hierarchy of different nodes is defined. Each cluster has its cluster head (CH). Sensor nodes in each cluster get the information and send it to CH. CHs collect the information from sensor nodes in its cluster region, aggregate the collected information, and send it directly to BS or next hop according to the predefined algorithm working in CH.

Low energy adaptive clustering hierarchy (LEACH) protocol is a base for the development of many hierarchical routing protocols in WSN. It is adaptive clustering and self-organizing routing protocol [1]. LEACH distributes the deployed nodes area into number of clusters. In each cluster, one node act as a CH and remaining nodes in this cluster act as a member of the cluster. These member nodes only communicate with their CH and CHs communicate with sink node or base station (BS) [1, 23]. CH is used as an intermediate node for member nodes to reach BS. CH collects data from member nodes, aggregates it, and forwards compressed data to BS. Due to added tasks, CH consumes more energy as compared to the normal nodes. As in static clustering, CH remains permanently, which results in the quick death of CHs [1, 24].

LEACH noticeably improves network lifetime and minimized energy dissipation as compared to other nonhierarchical routing protocols. But there are a lot of opportunities to enhance the capabilities of the LEACH protocol. LEACH is not suitable for large-scale networks, because of its single hop routing operation, irrespective of the distance; each CH has to communicate directly with the BS. At the time of selection of CHs, the residual energy of the node is not considered and it is the possibility that a node with less energy can be selected as CH; if that happened then it will become dead and consequently, that cluster becomes inaccessible. LEACH assumes even consumption of energy for every CH and does not guarantee proper CH distribution.

MODLEACH [21] is a modified version of LEACH protocol. In MODLEACH an efficient CH replacement scheme has been introduced. A predetermined threshold level has been set for the replacement of CHs. If current CH has enough battery power, which is greater than the predetermined threshold level then it will continue to serve as CH for the next round. The CH does not change until its battery power becomes less than the threshold limit. By using this CHs selection technique, energy consumed in the routing of update packets for newer CHs has been saved. MODLEACHHT [21] and MODLEACHST [21] are extended versions of MODLEACH [21]. In these versions, the idea of the hard and soft threshold level is introduced [25]. The soft threshold level is a little variation in the value of a recognized attribute which elicits the node to turn on the aerial and pass on data. The hard threshold is the attribute outright value beyond which the node that recognized threshold value will activate its transmitter and connect to CH. MODLEACHST [21] and MODLEACHHT [21] adopted the reactive approach and produced comparatively better results than MODLEACH [21]. MODLEACH, MODLEACHST, and MODLEACHHT improve CH selection technique but still, there are weaknesses in routing technique. These routing protocols adopt single hop routing strategy and are not suitable for large-scale networks. MODLEACH works on the basis of the density of sensor nodes which can make it unstable during the setup phase.

Multihop LEACH (MH-LEACH) routing protocol [20] adopts a multihopping strategy to send collected data to BS. In it every sensor node sends the collected data to the CH, CHs perform aggregation operation and forward data

to next CH until it reached BS. An optimal path is adopted between the sensor node and BS. The major limitation in MH-LEACH is delay factor which is due to multihop during transmission of data. As a lot of hops are added to reach BS, a little bit energy efficiency is achieved, which can be improved by reduction of unnecessary hops involved in the communication process.

Assisted LEACH (A-LEACH) [26] introduces additional node for load sharing of CHs, which is known as helper node. In every cluster along with CH, a helper node is also selected. A node which has sufficient remaining energy and is nearest to the BS is selected as helper node. Every CH receives the data sensed by the sensor node in each cluster. CHs forward the collected information to helper node after performing aggregation and removal of redundant data. The helper node performs routing tasks, which forwards the data to the nearest helper node. During routing phase, only helper nodes remain active and all other sensor nodes including CHs will go into sleeping mode and minimize energy dissipation.

Advanced Zonal Rectangular LEACH (AZR-LEACH) [27] have enhanced CH selection technique and introduced static clustering technique. The network deployment area is distributed into three logical partitions, such as advanced clusters, rectangular clusters, and zones. The rectangular clusters are formed by dividing the entire network into fixed clusters. Normally BS is installed at the center of the network area. The clusters those are around BS are advanced clusters and nodes under advanced clusters are considered as advanced nodes. Advanced CHs receive data from their member nodes as well as from other CHs and forward it to the BS. As advanced clusters are closer to BS, they consume less transmission power in comparison to other CHs. The zone is formed by the group of rectangular clusters. It is necessary that every zone must contain minimum one advanced cluster. AZR-LEACH [27] adopts a different strategy for the selection of CHs as compared to LEACH routing protocol. The node with highest remaining energy in the rectangular cluster will be elected as the CH. All nodes in the cluster send their remaining energy information to the CH. The main advantage of AZR-LEACH [27] is that the whole network area equally distributed into subareas, which equalize the network traffic load.

Centralized Low Energy Adaptive Clustering Hierarchy (LEACH-C) [2, 28] protocol introduces centralized cluster creation technique. Apart from CH's selection, all other operations are similar to LEACH routing protocol. Steady-state phase is different from original LEACH and setup phase is similar to LEACH protocol. In LEACH-C, all nodes send their location information and remaining energy level to the BS. For location information, sensor nodes are equipped with GPS module or any other tracking system. This information is shared at the beginning of each round. When BS has all the required information of nodes in the network then BS determines the value of average energy of all the sensor nodes in the network. The nodes with more remaining energy than the calculated average energy will be marked as candidate nodes. From the group of nodes which are marked as a candidate, BS will select a group of CHs using the simulated annealing. After choosing the selected group of CHs, it will be broadcasted to the entire network. A deterministic threshold algorithm is used by LEACH-C to collect the information of remaining energy level in the sensor nodes and to keep a record of nodes, those were selected as CHs in the previous rounds. By keeping CH's selection task centralized, this technique improves the energy efficiency and reduced the load on CHs but there is extra overhead on the BS. The performance of LEACH-C [28] diminishes when energy utilization for communication with BS increases, then the energy cost for cluster formation.

In order to prolong the stability time period (the time period before the first sensor node becomes dead), a two-level heterogeneous routing protocol is introduced, which is called stable election protocol (SEP) [29]. SEP distributes sensor nodes into two categories: normal sensor nodes and advanced sensor nodes. Advance sensor nodes are special nodes which have more energy (battery power) than normal sensor nodes. Both normal and advanced sensor nodes use weighted probability for the selection of CHs. As compared to advance sensor nodes, normal sensor nodes have lesser chances to become CH. Stability period is critical for most of the applications where reliable feedback is required from the sensor network. SEP routing protocol has noticeably improved the stability period than LEACH routing protocol. The major shortcoming in SEP routing protocol is that effective client deployment of sensor nodes is not guaranteed.

Enhanced stable election protocol (E-SEP) [30] has introduced three-level communication hierarchy. E-SEP distributes sensor nodes into three categories: normal sensor nodes, intermediate sensor nodes, and advanced sensor nodes, where intermediate sensor nodes have more energy (battery power) than normal sensor nodes and advance sensor nodes have more energy (battery power) than normal nodes and intermediate nodes. By using an extra level of heterogeneity as compared to SEP [29], up to some extent energy dissipation is reduced. Multihop routing with stable election protocol (MR-SEP) [31] enhances SEP routing protocol by dividing the network field into multiple layers of clusters. In each layer, CHs are selected and member sensor nodes join CHs in each layer according to their distance from CHs. For the transmission of data, CHs in each layer collect data from member nodes and collaborate with the CHs of adjacent layers. The CHs in upper layer perform as super CHs for the CHs of the lower layer. By adopting multiple layering approaches CHs are evenly distributed in the network field and multihoping strategy has increased the stability period but did not get any convincing improvement in the overall network lifetime.

3. Proposed Solution

In the development of routing protocol, the operating environment is provided by the network model, which consists of N sensor nodes, forwarder node, and BS. Sensor nodes are deployed randomly in the network area, and forwarder node is deployed in the network area, where it may be maximum involved in the communication process and BS is located outside of the network area. The major properties of the network model are as follows:

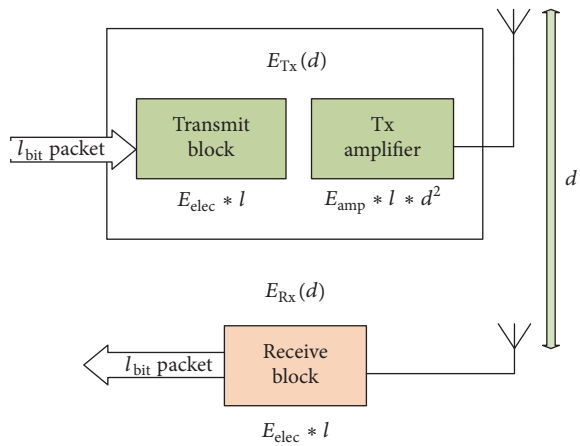


Figure 1: Radio communication model.

(i) All sensor nodes are homogeneous, having similar communication, sensing, and processing capabilities.

(ii) The sensor nodes are energy constrained.

(iii) In order to vary the transmission power of the sensor nodes, these are equipped with power control capabilities. The transmission range can be varied on the basis of requirements.

(iv) The forwarder node has more power and its battery can also be replaced and recharged.

3.1. Energy Model. It is the fact that the energy available to the sensor nodes is not only limited that it may diminish very easily if it is not properly managed. The main reasons for the consumption of energy in wireless sensor networks are communication and processing, with communication being the main responsible for the consumption of energy. First-order radio model [1, 2] is used for the energy model of sensor nodes. As shown in Figure 1, energy model consists of three main modules: receiver, transmitter, and power amplifier. The receiver consumes energy to run the receiver circuitry at the time of reception of data, and the transmitter consumes energy to run the power amplifier circuitry and transmitter circuitry at the time of transmission of data. Energy dissipation for transmitter and receiver is represented by E_{elec} and energy dissipation for transmit amplifier is represented by E_{amp}.

There are two transmission models: transmission model for free space and two-ray ground [32, 33]. In free space transmission, there is a direct line of sight path between the transmitting and the receiving nodes. In two-ray ground transmission model, the transmission between transmitting and the receiving nodes is not direct and the electromagnetic wave reached at the receiver from different paths at different times. The energy consumed for the transmission of l bits data packet, with distance "d" and energy consumed for the reception of "l" bits data by the receiver nodes, is represented by

$$E_{Tx}(l,d) = E_{elec} \times l + E_{amp} \times l \times d^2$$

$$E_{Rx}(l) = E_{elec} \times l. \tag{1}$$

Two different levels of power amplification for communication signals are introduced by MODLEACH [21], which is used on the basis of transmission nature. Equations (2) show the amplification levels for different communications on the basis of type and distance between communication devices. As compare to BS to cluster transmission, for intracluster transmission, lesser energy amplification level is used. Apart from energy saving benefits, collisions are reduced by the use of multiple power levels, and the number of packet drops and interference with other signals are also reduced.

$$\begin{aligned}
&\text{Energy Amplification level (CH to BS/Forwarder) } d \\
&\quad \geq d_0\left(E_{afs}\right) = \frac{10\,\text{pJ}}{\text{bit}}/\text{m}^2 \\
&\text{Energy Amplification level (CH to BS/Forwarder) } d \\
&\quad \leq d_0\left(E_{amp}\right) = \frac{0.0013\,\text{pJ}}{\text{bit}}/\text{m}^2 \\
&\text{Energy Amplification level (Intra Cluster Com.) } d \\
&\quad \geq d_1\left(E_{afs1}\right) = \frac{E_{afs}}{10} \\
&\text{Energy Amplification level (Intra Cluster Com.) } d \\
&\quad \leq d_1\left(E_{amp1}\right) = \frac{E_{amp}}{10}.
\end{aligned} \tag{2}$$

Transmission within a cluster is called intracluster communication. In it, sensor nodes sense data from the environment and send the collected data to the CH. Minimum amplification energy is required for intracluster communication.

3.2. Design of Energy-Efficient Multistage Routing Protocol (EE-MRP). The fundamental theme of our proposed routing protocol is the energy efficiency in larger network field of wireless sensor networks, where data is extremely interrelated and the requirement of end user is an only high-level function of data that contains a collection of events collected from the environment. The clustering hierarchical approach, efficient CH selection algorithm, and optimized routing algorithm are essential to design efficient solution for larger scale networks [33, 34]. The architectural design of our proposed routing protocol is shown in Figure 2. In our proposed routing protocol, homogeneous sensor nodes are randomly placed in the network field. Forwarder node is placed in the field where it may be maximum involved in the communication process. BS is placed outside of the network field.

3.2.1. Setup Phase. In this phase, the network field is divided into three logical stages (S1, S2, and S3) on the basis of sensor nodes located in the network field. BS is responsible for the division of network field into three logical stages. S1 and S3 have clustered regions and S2 is nonclustered region. Sensor nodes in S1 send data to the CHs, aggregation is done by CHs, and then aggregated data is forwarded to BS. Sensor

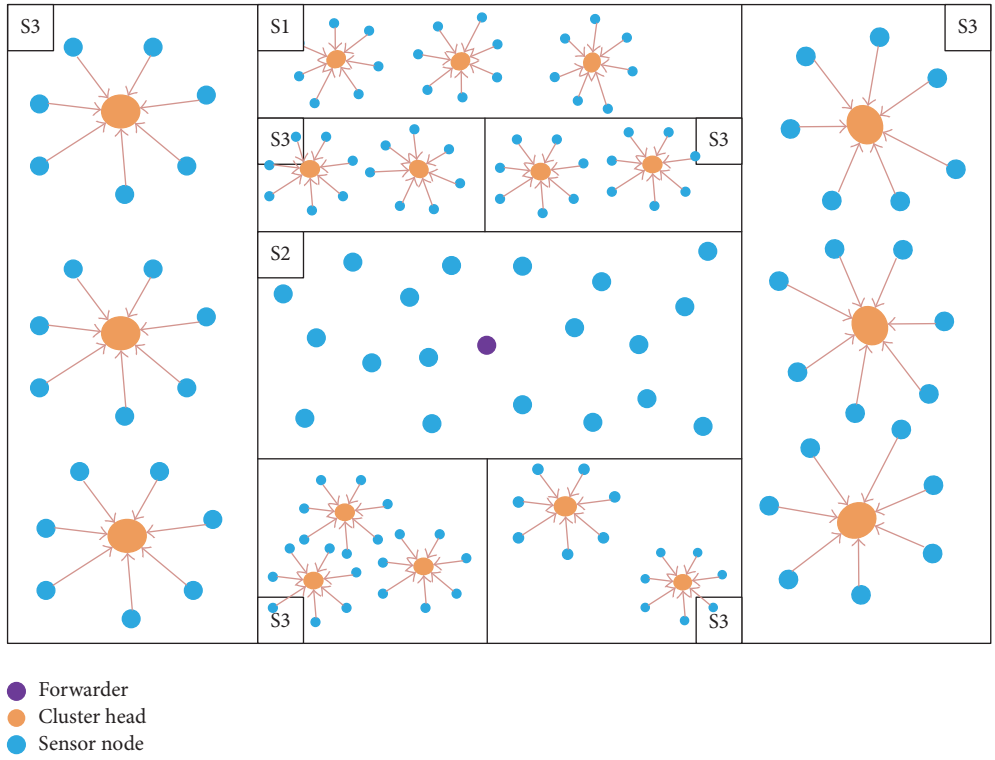


Figure 2: Basic architecture of EE-MRP.

nodes in S2 send data to the forwarder node, which performs aggregation on the collected data and then aggregated data is forwarded to BS. The sensor nodes which belonged to S3 also sends collected data to CHs and after aggregation, and CHs send to the forwarder node. The forwarder node then sends the collected data to the BS. By adding the forwarder node in the communication infrastructure, the distance between communicating nodes becomes less and energy consumption becomes less as compared to direct communication between CHs and BS.

3.2.2. Cluster Formation. In most of the cluster-based routing protocols, clusters are formed in the network field and CHs are randomly divided in the entire network field. In it, there are chances that CHs are not uniformly distributed in the network field. Some portion of the network may get more CHs and some portion may get lesser CHs. Due to this problem sensor nodes in that portion of network field may expire earlier and a part of the network becomes isolated. In EE-MRP initially, BS divides the network field into multiple logical segments. Then in each segment, CHs are selected. By using this mechanism CHs are evenly distributed in each segment of the network field. This cluster formation strategy increased the overall lifetime of the network. On the basis of location information, segment identification numbers are allotted to sensor nodes; therefore, sensor nodes can only join CHs located in their own segment. The cluster formation flow diagram is shown in Figure 3. Once CHs are selected, then CHs broadcast the invitation message to normal nodes in its area to join CH and become member node. Normal sensor nodes wait for invitation message from CHs in their area, and sensor nodes send join message to the nearest CH, on the basis of distance information.

3.2.3. Cluster Head Selection. Later to cluster formation, each node takes a decision whether or not to serve as a CH for the existing round. Every sensor node elects itself as a CH on the basis of the desired ratio of CHs and the status of eligibility flag to become CH. For instance, node n chooses a random number ranging from 0 to 1. The node will become CH if the threshold $T(n)$ is greater than a number. Following formula is used for the calculation of $T(n)$ [1, 11, 21].

$$T(n) = f(x) = \begin{cases} \dfrac{P}{1 - P * (r \bmod (1/P))}, & \text{if } n \in G, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where

n is total number of sensor nodes.

P is preferred the percentage of CH.

r is current round.

G is set of sensor nodes eligible to become CH.

Selected CHs broadcast their status to other nodes via MAC protocol, that is, Carrier Sense Multiple Access (CSMA). Member nodes compute Received Signal Strength Indication (RSSI) for the selection of CHs. Time Division Multiple Access (TDMA) schedules are formed by CH for accompanying member nodes in the cluster. Member nodes

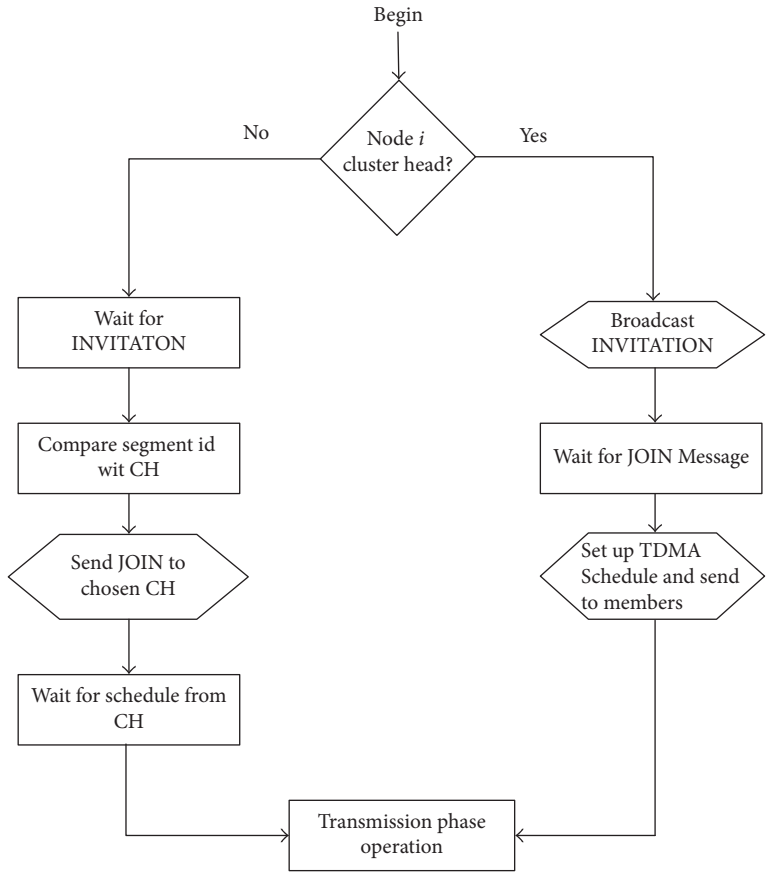


Figure 3: Cluster formation flowchart.

communicate with CH in the allotted time slots and remain in sleeping mode during unallocated time slots.

3.2.4. Cluster Head Replacement Scheme. In most of the hierarchical routing protocols like LEACH, the cluster changes in every round, and once a sensor node is selected as CH, it will never get another chance to become CH in the upcoming $1/p$ rounds. In every round, CHs are selected and the process for the formation of the cluster is repeated. In EE-MRP an efficient CH replacement scheme is adopted. The flow diagram of CHs replacement scheme is shown in Figure 4. In it a threshold level is used for the replacement of CHs, if the energy level of existing CHs is more than the predefined threshold then existing CH will remain until it crosses the threshold limit. When CH energy becomes lower than the threshold limit, then the availability flag for that sensor node is set to "0," which means that this sensor node is not available to be elected as CH; this ensures that any retired CH may not trigger the next round of CH change. The CH selection range is limited to the smallest range which applies to this very CH and its members. By using this efficient CHs replacement mechanism, unnecessary energy usage in the process of routing packets for new CHs and for cluster formation can be avoided.

3.2.5. Steady-State Phase. The communication paradigm is shown in Figure 5. After selection of CHs and allocation of TDMA slots, the process of steady-state phase begins. On the basis of TDMA protocol, communication is started between the sensor node and their respective CH, in their predefined allocated time slots. During the unallocated time slots, sensor nodes remain in sleep mode. By using this approach, better energy efficiency is achieved. The CHs perform aggregation on the collected data and forward it to the forwarder node or BS according to its segment number. Forwarder node also performs aggregation on the collected data and then transmits to the BS. Figure 6 shows the flowchart of the routing algorithm.

The communication algorithm works on the basis of three stages {S1, S2 and S3}. The sensor nodes belong to stage S1 and send the collected data to the CHs and CHs perform aggregation on the collected data and then send it to the BS. The sensor nodes belong to stage S2 and send the collected data directly to forwarder node and after aggregation forwarder node send the collected data to the BS. The sensor nodes belong to stage S3 and send the collected data to CHs after aggregation CHs send collected data to forwarder node and after aggregation forwarder node sends the collected data to BS.

4. Performance Evaluation

Simulations are done using MATLAB R2013b (8.2.0.701). MATLAB provides an interactive environment for the

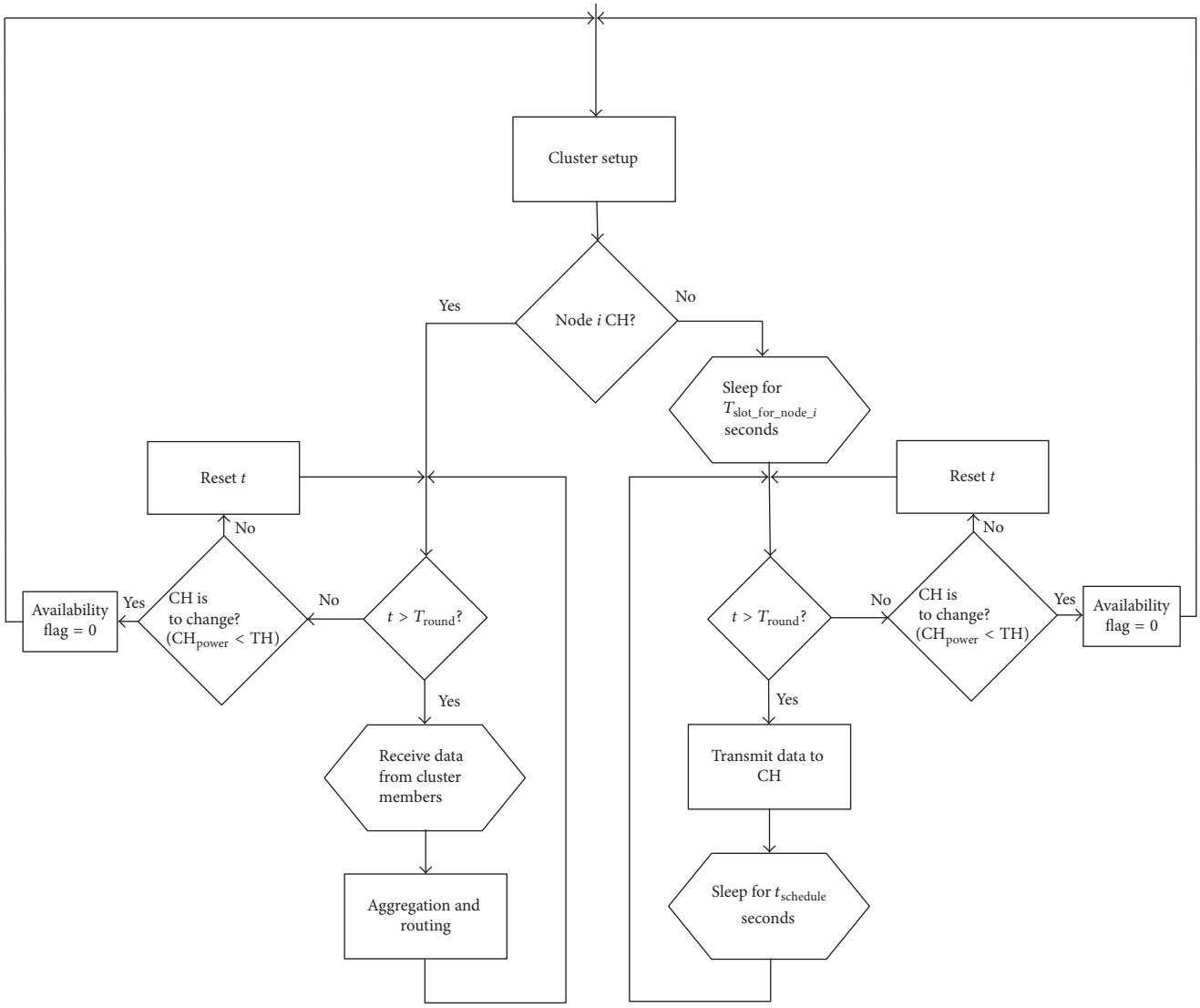


Figure 4: Cluster head replacement flowchart.

deployment of algorithms, virtualization of data, analysis of data, and numeric computation using a high-level technical computing language. In order to simulate our proposed routing protocol, homogeneous sensor nodes are randomly deployed. The wireless communication channel which is used between the sensor nodes deployed in the network depends on the transmission range and distance.

4.1. Network Scenario Assumptions and Parameters. It is assumed in the simulation that sensor nodes are densely and randomly scattered in a two-dimensional square field. Figure 7 shows a sample of randomly deployed sensor nodes used in our experiments. Sensor nodes are randomly deployed in the area of 150 m × 150 m. BS and forwarder are located at (75, 170) and (75, 65), respectively. Different samples of randomly deployed sensor nodes are used for each simulation and the results discussed in the next sections are the average values of 25 simulations.

The following properties are applied on the WSN.

(i) Sensor nodes are energy constrained, that is, not rechargeable, and always contain data to send.

(ii) There is only one BS, which is deployed outside of the network field.

(iii) There is only one forwarder in the field, which has a longer lifetime than normal sensor nodes and whose battery can be replaced.

(iv) A sensor node is declared as a dead node when it is unable to transmit data.

(v) It is assumed that the probability of interference and signal collision is ignorable.

The parameters utilized in the simulation are shown in Table 1.

4.1.1. Evaluation Parameters. The performance of hierarchical routing protocols for WSNs is evaluated on the basis of following terms.

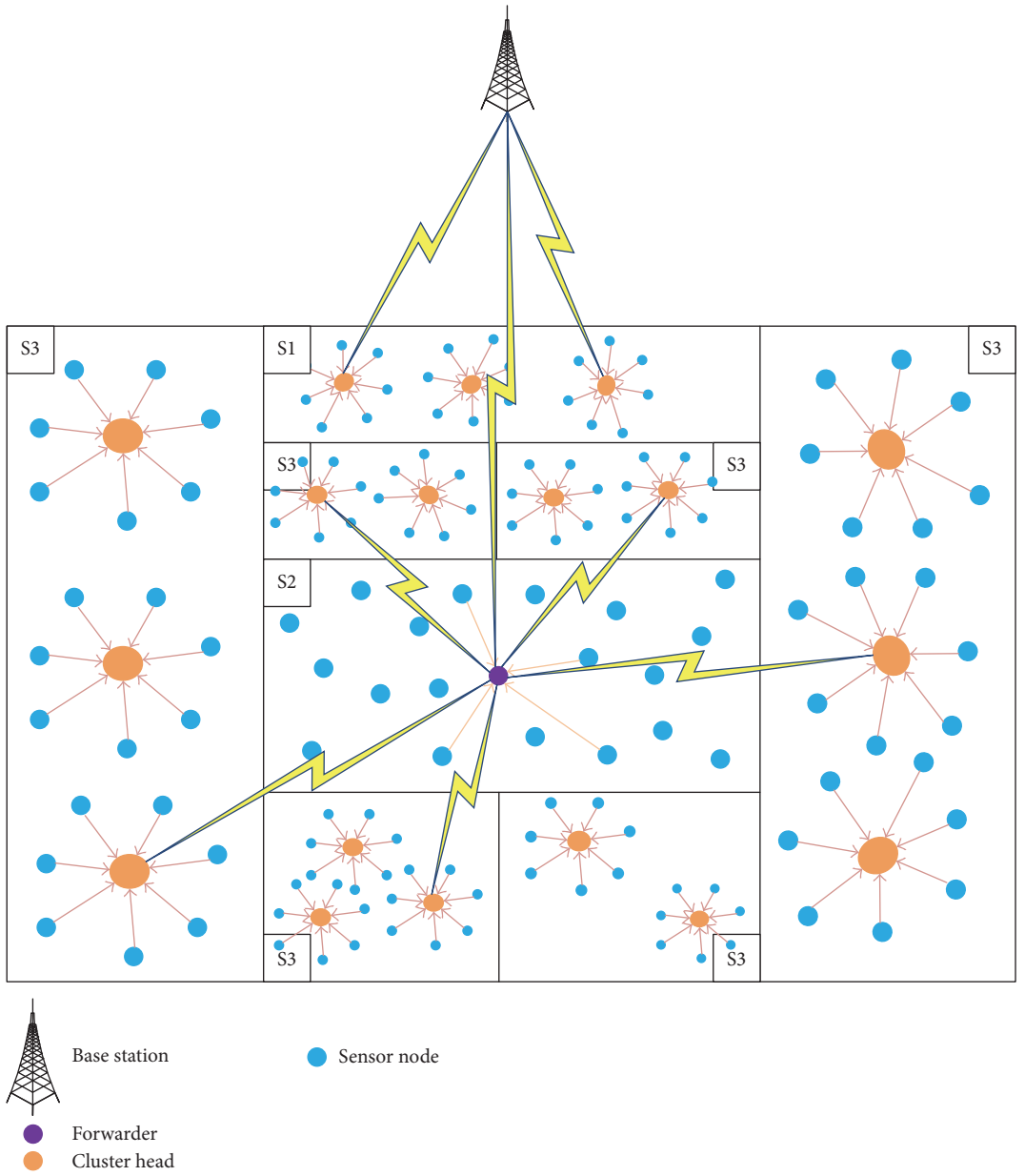


Figure 5: Communication paradigm.

Stability Period. It is the time period until the first sensor node becomes dead in the network field. The time interval between the startup of the WSNs operation and the time slot when first sensor node becomes dead is called stability period.

Instability Period. The time interval between time slot when first sensor node becomes dead and the time slot when last sensor node becomes dead is called instability period.

A Number of Alive Nodes. These are the total number of sensor nodes, those have not yet exhausted all of their energy and have enough energy to continue communication operation.

A Number of Dead Nodes. These are the total number of sensor nodes, those have exhausted all of their energy and did not have enough energy to continue communication operation.

Throughput. The rate of data sent from sensor nodes to CHs, from sensor nodes to forwarder, from CHs to forwarder, from CHs to BS, and from forwarder to BS, is collectively known as throughput.

Reliability. The comparison between stability period and instability period characterizes the strength of reliability. The longer stability period and shorter instability period show better reliability.

Overall Network Lifetime. The time period from the start of the network operation up to the death of the last sensor node is called overall network lifetime.

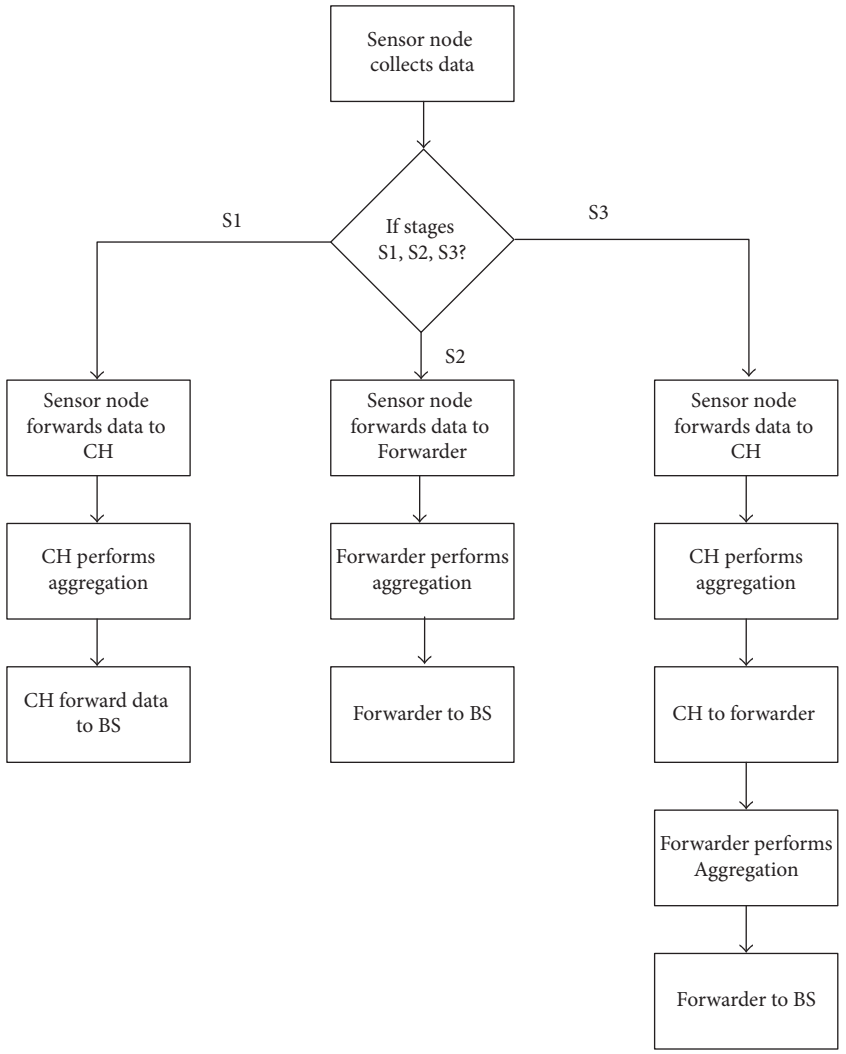


Figure 6: Flowchart of routing algorithm.

Table 1: Network parameters.

Parameters	Value
Network area (meter)	150×150
Number of nodes (N)	100
BS location	(75, 170)
Forwarder location	(75, 65)
Initial energy (E_0)	0.5 J
E_{TX}	50 nJ
E_{RX}	50 nJ
E_{amp} (cluster to BS/forwarder)	0.0013 pJ/bit/m^4
E_{fs} (cluster to BS/forwarder)	10 pJ/bit/m^2
E_{amp1} (intracluster comm.)	Eamp/10
E_{fs1} (intracluster comm.)	Efs/10
E_{da}	5 nJ/bit
Packet size	4000 bits
Number of rounds	3000

There is a trade-off between overall network lifetime and reliability. As overall network lifetime includes both stability period and instability period, therefore, WSN which has longer instability period is less stable but has longer overall network lifetime. On the other hand, WSN which has shorter instability period is more stable but has shorter overall network lifetime.

4.2. Simulation Results and Discussion. The simulation of our proposed routing protocol is done in comparison with LEACH [1] and MODLEACH [21], for the adherence of alive sensor nodes per round, dead sensor nodes per round, throughput, and overall network lifetime.

Figure 8 shows the number of alive nodes with respect to the number of rounds. It is shown that EE-MRP has comparatively more stability period than LEACH and MOD-LEACH. The first sensor node of EE-MRP becomes dead after approximately 1100 rounds while the first sensor node of LEACH and MODLEACH routing protocol becomes dead

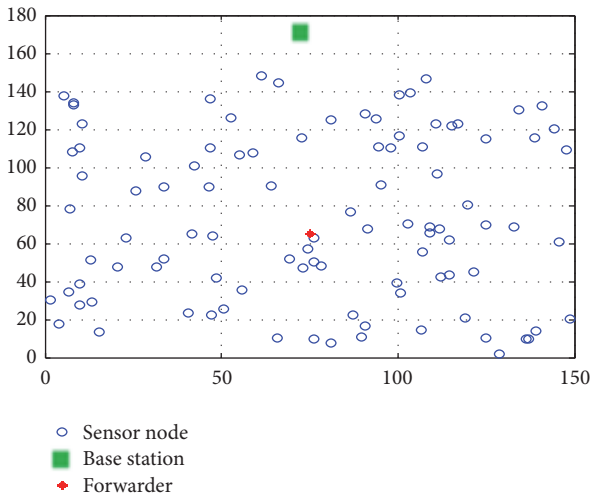


Figure 7: Random deployment of sensor nodes.

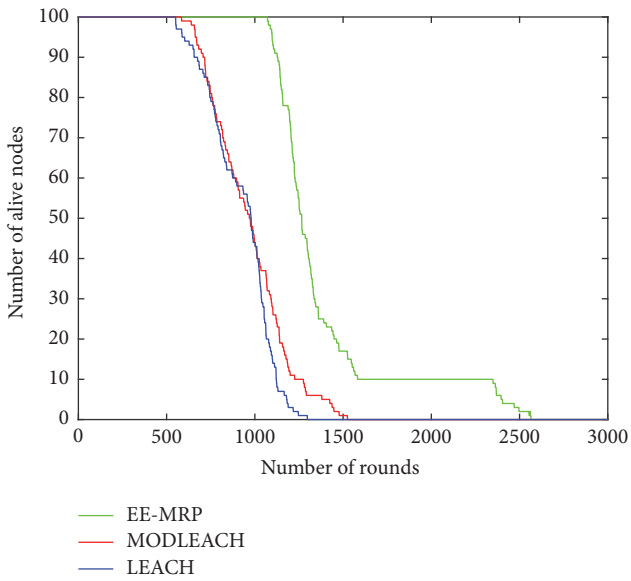


Figure 8: Total number of alive nodes in each round.

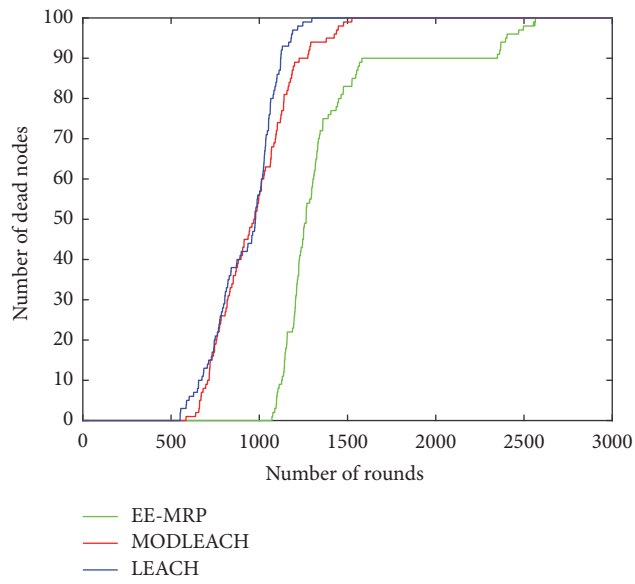


Figure 9: Total number of dead nodes in each round.

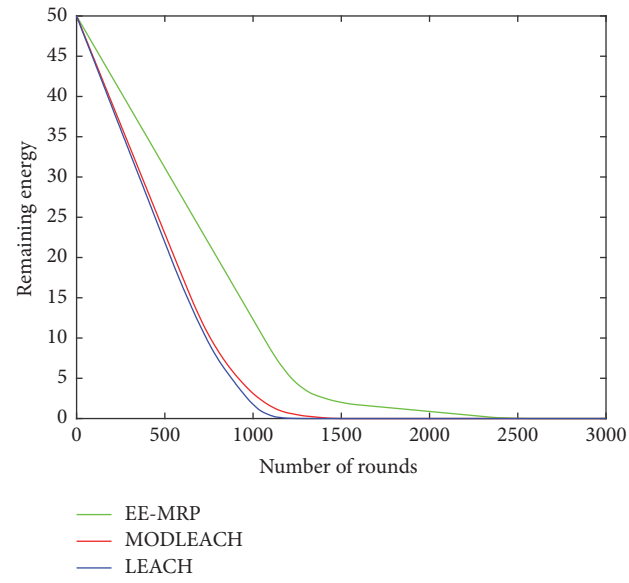


Figure 10: Analysis of remaining energy per round.

after approximately 580th round and 600th round, respectively. At 600th round, sensor nodes of both LEACH and MODLEACH routing protocol sharply start to become dead, and all sensor nodes of LEACH and MODLEACH become dead at 1300th round and 1500th round, respectively. The sensor nodes of EE-MRP start to become dead comparatively slowly after 1100th round and all sensor nodes become dead up to 2600th round.

The number of dead nodes with respect to the number of rounds is shown in Figure 9. It is observed that LEACH routing protocol has the shortest stability period and also has shorter instability period than other two routing protocols. The instability period of EE-MRP is longer but has much longer overall network lifetime than LEACH and MODLEACH. As the instability period of EE-MRP is longer than both LEACH and MODLEACH therefore EE-MRP has comparatively less reliability.

The average residual energy of network with respect to a number of rounds is shown in Figure 10. We have assumed that a sensor node has a maximum of 0.5-joule initial energy; therefore total energy for a network of 100 nodes is 50 joules. It is clearly observed in Figure 10 that energy dissipation of MODLEACH is less than LEACH. EE-MRP performs better in terms of energy dissipation per round and outperforms the LEACH and MODLEACH routing protocols.

Figure 11 shows the number of packets received by BS with respect to the number of rounds. It shows that throughput of EE-MRP is significantly greater as compared to LEACH and MODLEACH. From the graph, it is depicted that EE-MRP has better throughput up to 580% as compared to

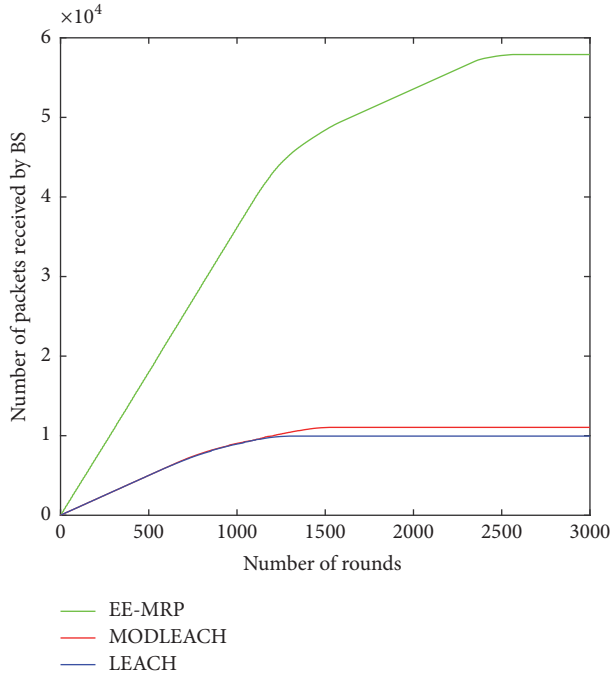


Figure 11: Comparative throughput.

LEACH routing protocol and up to 483% as compared to MODLEACH routing protocol.

The proposed routing protocol performed better than competing protocols in terms of various evaluation parameters discussed in Section 4.1.1. There are multiple reasons behind improved performance of proposed routing protocol. Firstly, efficient cluster head replacement mechanism avoids unnecessary repetition of CHs rotation. Secondly, use of forwarder node improves communication process. Thirdly, even distribution of CHs in the network field improves overall network lifetime and throughput. Fourthly, multistage transmission mechanism improves communication process and maximizes the stability period. All these factors contribute to the achievement of energy efficiency in our proposed routing protocol.

5. Conclusion and Future Work

The WSNs have many resource limitations; however, the energy limitation of the sensor nodes is one of the important characteristics among all due to its attachment to the life of the sensor nodes. Routing protocol plays an important role in optimizing energy consumption of the sensor nodes and in the maximization of the overall network lifetime. Therefore, efficient utilization of the available resources is the primary concern in the development of routing protocols for WSNs. We have developed an efficient clustering-based energy-efficient multistage routing protocol that meets the challenges of energy in WSNs. By dividing the network field into multiple stages, the CHs are evenly distributed which increases the throughput of the network and increases the overall lifetime of the network. Unnecessary rotation of CHs is avoided by adopting threshold based CHs selection mechanism. In order to minimize the distance between communicating nodes and CHs, the concept of forwarder node is introduced, which minimizes the routing distance between communicating nodes, CHs and BS. Multiple power amplification levels are used for intracluster communication and for communication between CHs and forwarder or BS. As less amplification power level is required for the intracluster transmission, by adopting multiple amplification power levels, unnecessary consumption of energy is avoided.

The performance of EE-MRP is evaluated using MATLAB simulation tool. Energy efficiency, throughput, and network lifetime are described as the performance metrics, used for comparison between our proposed routing protocol (EE-MRP) and existing routing protocols (LEACH and MOD-LEACH). It is clearly shown in the simulation results that EE-MRP surpassed the existing routing protocols in most of the performance metrics. In addition, with the help of these results, it has been verified that EE-MRP have adopted efficient CHs selection scheme and by using forwarder based routing strategy, overall lifetime of WSN has been improved significantly. This research work has opened numerous exigent research directions, which can be further explored. The proposed solutions have mostly addressed the energy efficiency in routing protocol, which can be further extended for the improvement of energy efficiency in MAC layer. The energy efficiency of the routing protocol can be further increased by making it application-specific like temperature monitoring and using the threshold level for the transmission of data between sensor nodes and CHs, CHs to BS, and CHs to forwarder node, which minimizes the communication for data transmission, and energy consumption during communication can be saved. This research work may further be extended by modeling and implementation of QoS in WSNs [35, 36].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to acknowledge Cihan University-Erbil for their support and sponsorship of this project.

References

[1] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Siences (HICSS '00)*, vol. 2, IEEE, January 2000.

[2] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[3] K. Sarammal and R. A. Roseline, "AReview: wireless sensor networks and its application, platforms, standards and

tools," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 4, no. 8, pp. 2905–2911, 2013.

[4] H. Kour, "Hierarchical Routing Protocols in Wireless Sensor Networks," *International Journal of Information Technology and Knowledge Management*, vol. 6, no. 1, pp. 47–52, 2012.

[5] S. Md Zin, N. Badrul Anuar, M. Laiha Mat Kiah, and A.-S. Khan Pathan, "Routing protocol design for secure WSN: review and open research issues," *Journal of Network and Computer Applications*, vol. 41, no. 1, pp. 517–530, 2014.

[6] M. A. Mahmood, W. K. G. Seah, and I. Welch, "Reliability in wireless sensor networks: a survey and challenges ahead," *Computer Networks*, vol. 79, pp. 166–187, 2015.

[7] P. Kuila and P. K. Jana, "Energy efficient clustering and routing algorithms for wireless sensor networks: particle swarm optimization approach," *Engineering Applications of Artificial Intelligence*, vol. 33, pp. 127–140, 2014.

[8] S. R. BoselinPrabhu, E. Gajendran, and N. Balakumar, "Contemporary challenges in environmental monitoring application of wireless sensors," *International Journal of Universal Science and Engineering*, no. 2, 2016.

[9] A. M. Khedr and W. Osamy, "Minimum perimeter coverage of query regions in a heterogeneous wireless sensor network," *Information Sciences*, vol. 181, no. 15, pp. 3130–3142, 2011.

[10] G. Han, J. Jiang, L. Shu, J. Niu, and H. Chao, "Management and applications of trust in wireless sensor networks: a survey," *Journal of Computer and System Sciences*, vol. 80, no. 3, pp. 602–617, 2014.

[11] X. H. Wu and S. Wang, "Performance comparison of LEACH and LEACH-C protocols by NS2," in *Proceedings of the 9th International Symposium on Distributed Computing and Applications to Business Engineering and Science (DCABES '10)*, pp. 254–258, Hong Kong, China, August 2010.

[12] S. Gupta and K. C. Roy, "Comparison of different energy minimization techniques in wireless sensor network," *International Journal of Computer Applications*, vol. 75, no. 18, pp. 20–26, 2013.

[13] D. Zhang, G. Li, K. Zheng, X. Ming, and Z.-H. Pan, "An energy-balanced routing method based on forward-aware factor for wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 766–773, 2014.

[14] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 551–591, 2013.

[15] P. G. V. Naranjo, M. Shojafar, H. Mostafaei, Z. Pooranian, and E. Baccarelli, "P-SEP: a prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks," *The Journal of Supercomputing*, vol. 73, no. 2, pp. 733–755, 2017.

[16] H. Mostafaei, A. Montieri, V. Persico, and A. Pescapé, "A sleep scheduling approach based on learning automata for WSN partial coverage," *Journal of Network and Computer Applications*, vol. 80, pp. 67–78, 2017.

[17] H. Mostafaei, "Stochastic barrier coverage in wireless sensor networks based on distributed learning automata," *Computer Communications*, vol. 55, pp. 51–61, 2015.

[18] A. A. Aziz, Y. A. Sekercoioglu, P. Fitzpatrick, and M. Ivanovich, "A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 121–144, 2013.

[19] M. Samanta and I. Banerjee, "Optimal load distribution of cluster head in fault-tolerant wireless sensor network," in *Proceedings of the 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS '14)*, March 2014.

[20] V. Biradar, S. R. Sawant, R. R. Mudholkar, and V. C. Patil, "Multihop routing in self-organizing wireless sensor networks," *International Journal of Computer Science Issues (IJCSI)*, vol. 8, no. 1, 2011.

[21] D. Mahmood, N. Javaid, S. Mahmood, S. Qureshi, A. M. Memon, and T. Zaman, "MODLEACH: a variant of LEACH for WSNs," in *Proceedings of the IEEE 8th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA '13)*, pp. 158–163, Compeigne, France, October 2013.

[22] A. Kaur and A. Grover, "LEACH and extended LEACH protocols in wireless sensor network-a survey," *International Journal of Computer Applications*, vol. 116, no. 10, pp. 1–5, 2015.

[23] M. Sharma and K. Sharma, "An energy efficient extended LEACH (EEE LEACH)," in *Proceedings of the International Conference on Communication Systems and Network Technologies (CSNT '12)*, pp. 377–382, May 2012.

[24] R. M. Bani Hani and A. A. Ijjeh, "A survey on LEACH-based energy aware protocols for wireless sensor networks," *Journal of Communications*, vol. 8, no. 3, pp. 192–205, 2013.

[25] A. Gupta and A. Nayyar, "A Comprehensive Review of Cluster-Based Energy Efficient Routing Protocols in Wireless Sensor Networks," *International Journal of Research in Computer and Communication Technology*, vol. 3, no. 1, 2014.

[26] S. Vinodh Kumar and A. Pal, "Assisted-Leach (A-Leach) energy efficient routing protocol for wireless sensor networks," *International Journal of Computer and Communication Engineering*, vol. 2, no. 4, pp. 420–424, 2013.

[27] Z. A. Khan and S. Sampalli, "AZR-LEACH: an energy efficient routing protocol for wireless sensor networks," *International Journal of Communications, Network and System Sciences*, vol. 05, no. 11, pp. 785–795, 2012.

[28] S. Shi, X. Liu, and X. Gu, "An energy-efficiency Optimized LEACH-C for wireless sensor networks," in *Proceedings of the 2012 7th International ICST Conference on Communications and Networking in China (CHINACOM '12)*, pp. 487–492, August 2012.

[29] G. Smaragdakis, I. Matta, and A. Bestavros, "Ibrahim Matta and AzerBestavros, SEP: A Stable Election for clustered heterogeneous wireless sensor network," Tech. Rep., Boston University Computer Science Department, Boston, Mass, USA, 2004.

[30] F. A. Aderohunmu and J. D. Deng, *An Enhanced Stable Election Protocol (SEP) for Clustered Heterogeneous WSN*, Department of Information Science, University of Otago, New Zealand, 2009.

[31] H. Kaur, H. Sharma, and G. Manu, "Multi-hop Routing SEP (MR-SEP) for clustering in wireless sensor network," *International Journal of Engineering Technology, Management and Applied Sciences*, vol. 2, no. 3, 2014.

[32] T. S. Rappaport, Ed., *Wireless, Communications; Principles and Practice*, Prentice Hall, PTR, New Jersey, NJ, USA, 2nd edition, 2001.

[33] A. Osseiran, J. F. Monserrat, and Y. P. Marsch, *5G Mobile and Wireless Communications Technology*, Cambridge University Press, Cambridge, UK, 2016.

[34] M. Hammoudeh and R. Newman, "Adaptive routing in wireless sensor networks: QoS optimisation for enhanced application performance," *Information Fusion*, vol. 22, pp. 3–15, 2015.

[35] P. K. Swain and M. K. Rath, "Analysis and QoS measurement for voluminous data flow in Wireless Sensor Network," in *Proceedings of the 3rd International Conference on Recent Advances in Information Technology (RAIT '16)*, pp. 61–64, March 2016.

[36] R. Fang, J. Wang, W. Sun, and Q. Li, "QoS Model of WSNs Communication in Smart Distribution Grid," *International Journal of Distributed Sensor Networks*, vol. 2016, Article ID 6926793, 23 pages, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 3279452, 14 pages
https://doi.org/10.1155/2017/3279452

Research Article

Adaptive and Blind Wideband Spectrum Sensing Scheme Using Singular Value Decomposition

Zhuhua Hu,[1,2] Yong Bai,[1,2] Yaochi Zhao,[1] Chong Shen,[1,2] and Mingshan Xie[1]

[1] *College of Information Science & Technology, Hainan University, Haikou, China*
[2] *State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University, Haikou, China*

Correspondence should be addressed to Yong Bai; bai@hainu.edu.cn

Received 15 August 2017; Revised 12 November 2017; Accepted 7 December 2017; Published 25 December 2017

Academic Editor: Waleed Ejaz

The Modulated Wideband Converter (MWC) can provide a sub-Nyquist sampling for continuous analog signal and reconstruct the spectral support. However, the existing reconstruction algorithms need a priori information of sparsity order, are not self-adaptive for SNR, and are not fault tolerant enough. These problems affect the reconstruction performance in practical sensing scenarios. In this paper, an Adaptive and Blind Reduced MMV (Multiple Measurement Vectors) Boost (ABRMB) scheme based on singular value decomposition (SVD) for wideband spectrum sensing is proposed. Firstly, the characteristics of singular values of signals are used to estimate the noise intensity and sparsity order, and an adaptive decision threshold can be determined. Secondly, optimal neighborhood selection strategy is employed to improve the fault tolerance in the solver of ABRMB. The experimental results demonstrate that, compared with ReMBo (Reduce MMV and Boost) and RPMB (Randomly Projecting MMV and Boost), ABRMB can significantly improve the success rate of reconstruction without the need to know noise intensity and sparsity order and can achieve high probability of reconstruction with fewer sampling channels, lower minimum sampling rate, and lower approximation error of the potential of spectral support.

1. Introduction

Spectrum resource has become increasingly scarce with emerging wireless services. Nevertheless, assigned radio spectrum to authorized users is mostly underutilized. Cognitive radio (CR) technology can improve frequency spectrum utilization by detecting and accessing the frequency range that has not been occupied by authorized users (primary users). Therefore, one of the crucial tasks in the cognitive radio system is to constantly monitor the potential spectrum bands and detect the activities of primary users [1, 2]. The methods of signal detection on narrow band mainly include energy detection [3], coherent detection [4], and feature detection [5]. Since the wideband spectrum sensing can provide more spectrum access opportunities for cognitive users (secondary users), multiband based wideband spectrum sensing has gained much more research attention in recent years [6, 7].

At the receiving side, the traditional method of acquiring RF (Radio Frequency) information is to use demodulation under the condition of known carrier. However, in practical scenarios, it is often required to directly perform blind sensing for high frequency wideband analog signals, which brings great challenges to spectrum detection. Firstly, extremely high sampling rate has exceeded the limit of physical ADC (Analog-to-Digital Converter) capability. Meanwhile, the storage and transmission for the sampled data can bring huge overhead. In addition, the carrier frequency of the received multiband RF signal is usually unknown. To address these challenges, compressed sensing theory is applied to wideband spectrum sensing based on compressed sampling and signal reconstruction [8, 9]. In order to perform sub-Nyquist sampling on continuous spectrum signal, Mishali et al. proposed the MWC (Modulated Wideband Converter) scheme which uses multichannel compressed sampling and reconstruction with parallel structure based on compressive sensing and time-frequency transform theory [10, 11]. In MWC system, the multiband analog sparse signal can be collected with sub-Nyquist sampling, and the spectral support of signal can be reconstructed by a CTF (continuous-to-finite) block.

From [10], it is known that the time domain reconstruction model of MWC can be attributed to the Multiple Measurement Vectors (MMV) problem. MWC has been considered to be an effective approach for compressed sampling and signal reconstruction with its applications on radar, broadband communication, and cognitive radio spectrum sensing [12–14].

The accurate reconstruction of spectral support of signal is the key to realize the spectrum sensing. At present, there is still much room for improvement in the success rate of sensing, the required minimum number of channels, and the maximum number of subbands that can be reconstructed. The solvability of MMV problems is important for MWC reconstruction ability. To enhance the solvability of MMV problem, Reduce MMV and Boost (ReMBo) algorithm, which can transform MMV problem into single measurement vectors (SMV) problem is proposed in [15]. In ReMBo, the sampling value matrix Y is projected on a random vector w which is independently and identically distributed in the interval $[-1, 1]$ and obtains a column vector v, $v = Yw$. Then, it tries to obtain the estimated spectral support of signal by iteratively computing SMV problem ($v = \Phi Zw = \Phi z$) for many times. Although ReMBo is better than most of the existing reconstruction algorithms in the success rate of reconstruction and computational overhead, there is still a big gap between the maximum number of reconstructed frequency subbands and the theoretical upper limit [16]. The total sampling rate and the minimum number of required channels of ReMBo algorithm are still much higher than the theoretical lower bound [10, 17]. To improve the performance of MWC system, a dimension adjustable Randomly Projecting MMV and Boost (RPMB) framework algorithm, which transforms the initial MMV problem into a series of low dimensional MMV problems with the same sparsity order, was proposed in [18]. In the solving process, the solution is tested several times until a satisfactory solution is obtained. Compared with ReMBo, RPMB can reduce the number of required hardware channels and improve the maximum number of reconstructed frequency subbands in the condition of precise reconstruction. In spite of this, the RPMB framework is still faced with difficult problems to be solved: (1) The reconstruction performance of RPMB is greatly reduced under low SNR because the influence of the noise intensity is not considered in the RPMB solver; (2) the RPMB is not flexible enough since the decision threshold of RPMB is a fixed value, which has no self-adaptive ability for the uncertainty of the noise; (3) RPMB needs to know the sparsity order of the signal in advance, which is difficult in some practical applications; (4) although the RPMB solver improves the algorithm's fault tolerance by finding more potential support bands, the effect of fault tolerance remains to be improved.

To address the above-mentioned problems, this paper proposes a Self-Adaptive and Blind Reduced MMV and Boost (ABRMB) scheme for MWC based on SVD [19]. Our main contributions are summarized as follows.

(1) ABRMB can first reveal the intrinsic properties of signal using SVD and perform the linear fitting on the noise singular value by using the tail singular values as noise data. Then, it can estimate the noise intensity, and the threshold of deciding support elements can be set adaptively.

(2) Our proposed scheme calculates the contribution of the estimated noise singular on the singular values of signal and figures out the sparsity order by using gradient and difference operation. Hence the number of signal subbands can be estimated.

(3) In the solver of ABRMB, the fault tolerance can be improved by the optimal neighborhood selection method. Compared with ReMBo and RPMB, when noise intensity and sparsity order are unknown, our proposed scheme can use fewer channels to achieve the high probability reconstruction of spectral support in low SNR range and increase the maximum subband number of reconstructed signal.

The rest of this paper is organized as follows. In Section 2, the related work is introduced. In Section 3, the multiband sparse signal model is introduced, and sampling principles of MWC are described as well as some of the present major issues. Section 4 introduces the key ideas of the noise intensity estimation, sparsity order estimation, and optimal neighborhood selection. In Section 5, the ABRMB scheme includes its solver which are specifically described, the convergence proof of which is also given. Section 6 performs validation and analysis on ABRMB from four important performance evaluation indexes. Conclusions and discussion are in Section 7.

The main notations used in this paper are listed in Notations to make them clearer and easier to read.

2. Related Work

Current research on the performance of spectrum sensing can be mainly classified into the following two categories. The first category, especially in the design of CRN, focuses on investigating the optimal trade-off between energy consumption and throughput for secondary users (SUs). The second category, especially for wideband spectrum sensing, mainly focuses on studying how to achieve a good balance between noise interference and sensing accuracy.

As for the first category, some earlier researches focus on the trade-off between the interference and throughput. Specifically, for the unslotted CRN, Yang et al. [20] designed an interference-constrained novel simultaneous sensing and transmission scheme. This scheme exploited the statistic information of the activities of primary users (PUs), and the transmission duration is adaptively adjusted to avoid the interference. In the same period, energy efficient techniques, in cooperative spectrum sensing (CSS), have also gained a lot of attention [6]. In [21], considering the energy consumption problems of each sensing node in cooperative CR sensor network, a user selection scheme was proposed to minimize the energy consumed on each CR node based on the binary knapsack problem and its dynamic solution. After that, Ejaz et al. [22] indicated that an optimal sensing, reporting, and transmission duration can be found while providing a given throughput and reliability constraint. The solution can obtain the best trade-off between energy consumption and throughput for SUs.

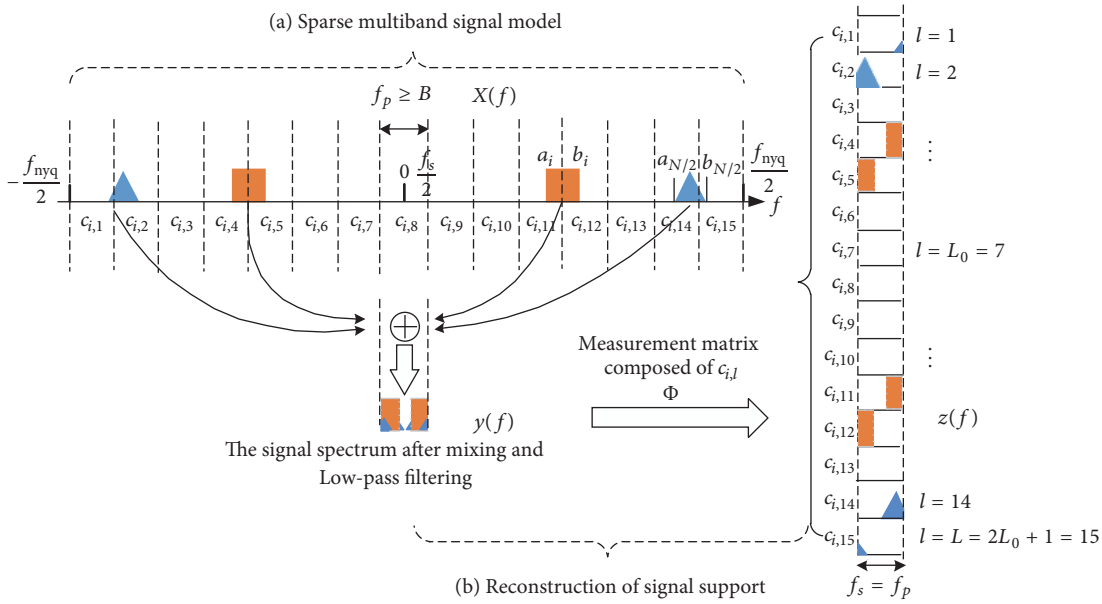


Figure 1: The reconstruction principles of the spectral support of MWC system.

As for the second category, in order to overcome the limitation of hardware technology and achieve wideband spectrum sensing, the research work mainly concentrated on how to get a good trade-off between the noise interference and sensing accuracy [4]. Firstly, based on the compressed sensing theory, a parallel MWC architecture for wideband spectrum sensing was designed by Mishali et al. [10, 11], which can achieve sub-Nyquist sampling for the wideband signal. Then, a MWC system with SwSOMP algorithm was proposed in [23]. Given the noise interference, the algorithm can get higher reconstruction precision of spectral support. In [24], in order to reduce the high hardware complexity of MWC, a compressive wideband spectrum sensing scheme based on a single channel was designed.

The work of this paper belongs to the second category. All of the aforementioned literatures only consider ideal noise interference model (NIM) and assume that the signal sparsity is known. However, in reality, NIM is always nonideal, where the intensity of noise varies with the change of communication environment. Moreover, the signal sparsity is always unknown in reality. In addition, we also hope to adopt a better reconstruction strategy and use fewer parallel channels to obtain higher reconstruction accuracy. Therefore, these problems need to be further investigated.

3. Problem Statement of MWC Model and Our Proposed Problem Solution

3.1. MWC System Model. The reconstruction principles of the spectral support of MWC system is shown in Figure 1. Figure 1(a) shows the sparse multiband analog signal model. Sparse multiband analog signal is common in cognitive radio [25]. Suppose the received signal $x(t)$ is a sparse bandpass analog signal. The spectrum is distributed in frequency interval $[-f_{nyq}/2, f_{nyq}/2]$, and f_{nyq} is Nyquist rate, which usually can reach GSPS order of magnitude. Suppose the spectrum of $x(t)$ only contains N subbands whose bandwidth is $B_i \leq B$ $(N \geq i > 0)$, and there is no overlap between subbands. B is the maximal bandwidth in subbands, and the center carrier frequency of each subband is unknown. P_N and B can be defined as

$$P_N = \bigcup_{i=1}^{N/2} \{(a_i, b_i) \cup (-b_i, -a_i)\},$$
$$B = \max_i (b_i - a_i). \quad (1)$$

In (1), P_N is union of all subbands whose amplitude is not zero, which represents the effective frequency component of $x(t)$.

As shown in Figure 1(a), the whole band is equally divided into L consecutive narrow band channels, every bandwidth of which is no less than B. Then the spectrum of $x(t)$ has at most $2N$ parts which get energy in the whole frequency band. If the channel is marked $[1, \ldots, L]$, the set of channel numbers corresponding to each subband $X_i(f)$ is the spectral support of $x(t)$, which is defined as $\Lambda = \text{supp}(X(f))$. $|\Lambda|$ is the number of nonzero elements in Λ, that is, the potential of Λ. The frequency bands corresponding to these numbers are called support bands. Since $2N$ is much smaller than L, $x(t)$ can be considered as sparse multiband signal. Figure 1(b) shows the reconstruction process of the support of $x(t)$. Assume that the number of bands is 4, $f_s = f_p$, and $f_p \geq B$. We divide the wideband spectrum into L spectrum slices, where $L = 2L_0 + 1$. In order to ensure that the discrete Fourier transformation result of sampling sequence contains all the components of the original signal spectrum $X(f)$, L_0 must satisfy $L_0 = \lfloor (f_{nyq} + f_s)/2f_p \rfloor - 1$. After mixing and low-pass filtering, the spectrum information of the original signal appears in the sampling interval $[-f_s/2, f_s/2]$, and the mixing

coefficient of each spectrum slice is c_{il}, where l is the index of spectrum slices. According to the theory of compressed sensing, we can obtain the spectrum support of the multiband signal.

In conclusion, the support bands of $x(t)$ must meet two conditions: (1) it is distributed in an extremely wide frequency range; (2) the support bands of signal only exist in a few discrete frequency spectrum.

3.2. Sampling in MWC System. MWC contains multiple parallel sampling channels, and each channel has same structure with mixer, low-pass filter, and ADC [10]. The received signal $x(t)$ is input to m parallel channels at the same time, and each channel is multiplied by periodic mixing signal $p_i(t)$ with different mode, which makes spectrum of $x(t)$ move to baseband. Each of the channels' $p_i(t)$ is not related. The period of $p_i(t)$ is $T_p = 1/f_p$, and M is used to show the number of random ±1 switches in a cycle. Mf_p is defined as the switching frequency of mixed signals. The mixed signals pass through the low-pass filter whose cut-off frequency is $1/2T_s$. At last, it passes through ADC whose sampling rate is $f_s = 1/T_s$ and obtains m groups low-speed digital sampling sequences $y_i[n]$.

On analysis of the ith channel, the Fourier series expansion of the random mixing function is

$$p_i(t) = \sum_{l=-\infty}^{\infty} c_{il} e^{j2\pi f_p l t}, \tag{2}$$

where $c_{il} = d_l \sum_{k=0}^{L-1} \alpha_{ik} e^{-j(2\pi/L)lk}$, $\alpha_{ik} \in \{-1, +1\}$. When $l = 0$, $d_0 = 1/L$, and when $l \neq 0$, $d_l = (1 - e^{-j(2\pi l/L)})/j2\pi l$. In (2), $p_i(t)$ denotes a pseudorandom sequence of ±1, which is used as a mixing signal of the ith sampling channel; l is index of spectrum slice; c_{il} is the coefficient of Fourier series expansion.

Next, after passing through the low-pass filter with frequency characteristic of $H(f) = \begin{cases} 1 & |f| \le f_s/2 \\ 0 & |f| > f_s/2 \end{cases}$, the relationship between DTFT (Discrete Time Fourier Transform) of $y_i[n]$ and the Fourier transform $X(f)$ of $x(t)$ is

$$Y_i\left(e^{j2\pi f T_s}\right) = \sum_{l=-L_0}^{L_0} c_{il} X\left(f - l f_p\right). \tag{3}$$

In (3), $f \in [-f_s/2, f_s/2]$, and L_0 is the smallest integer that can satisfy $L = 2L_0 + 1 \ge f_{\text{nyq}}/f_p$. Formula (3) shows that the spectrum of the output sequence $y_i[n]$ is equivalent to the shift weighted sum of the original signal spectrum $X(f)$ with f_p as its step, which is intercepted into spectral fragments with f_s width by low-pass filter. If $Y_i(e^{j2\pi f T_s})$ is considered as the ith component of m-dimensional column vector $y(f)$ and $X(f - lf_p)$ is considered as the lth component of $2L_0 + 1$-dimensional column vector $z(f)$, then (3) can be expressed as

$$y(f) = \Phi z(f), \quad f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]. \tag{4}$$

In (4), Φ is a $m \times L$ measurement matrix, $\Phi_{il} = c_{i,-l} = c_{il}^*$, $1 \le i \le m$, and $m < L$. If IDTFT (Inverse Discrete Time Fourier Transform) is performed on both ends of (4), we can get the corresponding relationship between the sequence $Z[n] = [z_1[n], z_2[n], \ldots, z_L[n]]^T$ and the sampling data $Y[n] = [y_1[n], y_2[n], \ldots, y_m[n]]^T$.

$$Y[n] = \Phi Z[n]. \tag{5}$$

For any frequency $f \in [-f_s/2, f_s/2]$, (5) is a typical compressed sensing problem. When sampled values matrix Y and an observation matrix Φ are known, a sparse vector Z can be recovered. Since MWC is a MMV problem, the one-dimensional measured value vector turns into a two-dimensional matrix. As a result, paper [15] gets a one-dimensional vector after projecting Y ingeniously and then reconstructs the signal support bands by using compressed sensing technique. Furthermore, the reconstruction by building a CTF (continuous-to-finite) module is described in [10].

3.3. Problem Statement. The principle of reconstructing spectral support for MWC is shown in Figure 1 where Figure 1(b) gives a specific reconstruction process. As can be seen in Figure 1, if the number of support bands of $x(t)$ is N, the maximum reconstructed subbands are $2N$ after passing through the MWC system; that is, the maximum possible sparsity order K is $2N$. In the MOMPMMV (Modified Orthogonal Matching Pursuit for MMV), the solver mentioned in paper [18], the initial solution $\ddot{Z}$ is obtained via the OMPMMV, which needs to know K beforehand. In addition, in the condition judgment of RPMB, the algorithm uses $|\widehat{\Lambda}| \le K$ to determine whether the potential of spectral support meets the requirements. However, in cognitive radio applications, it is difficult to know the sparsity order of signal. Therefore, aiming at the practical application, an effective method to estimate the sparsity order of signal is needed.

Secondly, in [18], the condition of deciding spectral support is $\|\widetilde{Z}^{i\rightarrow}\|_2 \ge \varepsilon$; and the condition of determining whether the spectral support meets the requirements in the iterative process of RPMB is $\|\widehat{Y} - \Phi\widehat{Z}\|_F \le \varepsilon$. As mentioned in Section 1, in the case where there is no noise or high SNR, ε can be figured by taking a fixed small value. If the SNR is low or fluctuates greatly, the value of ε will have to perform self-adaptive change according to the intensity of the noise. In paper [18], ε is a fixed value without considering SNR. Hence, RPMB algorithm is neither suitable nor flexible under low or fluctuated SNR. Thus an effective method is needed to estimate the intensity of the noise and adaptively change the decision threshold ε.

Finally, in MOMPMMV, to improve the fault tolerance ability, the method is $\ddot{\Lambda} = \max(\|\ddot{Z}^{i\rightarrow}\|_2, m)$, where m is the number of sampling channels, that is, taking m maximum norm 2 as the initial spectral support. This method is effective, but it is not the best way. For example, as shown in Figure 1(a), the $N/2$th subband $X_{N/2}(f)$ of $x(t)$ is located at the 14th and 15th narrow channels, and the spectrum energy of the 15th narrowband channel is quite small, as a result of which if the fault tolerance method in [18] is used, under low SNR, number 15 will not be able to merge into the initial support $\ddot{\Lambda}$. So it is necessary to design a better fault tolerance strategy.

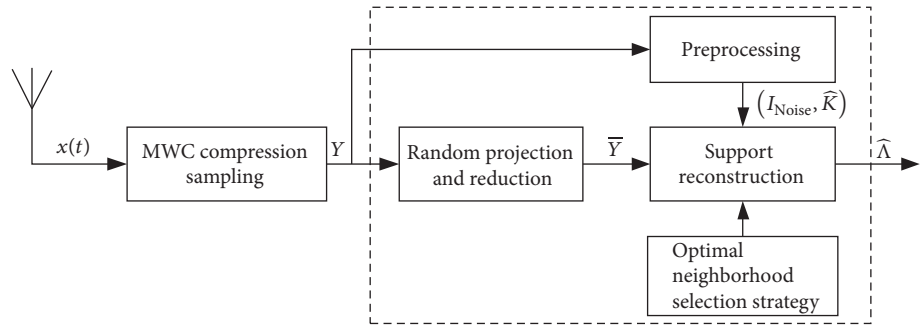


Figure 2: The flowchart of our proposed problem solution.

3.4. Brief Description of Our Proposed Problem Solution. Our proposed problem solution is shown in Figure 2 where multiband analog signal $x(t)$ passes through the MWC system to obtain compressed sampling data Y. Then, in order to reduce complexity, we reduce the dimension of Y by means of random projection. Subsequently, the signal sparsity and noise intensity are estimated in the preprocessing, namely, $\widehat{K}$ and I_{Noise}. Meanwhile, to improve the fault tolerance ability, we adopt the optimal neighborhood selection strategy, which is described in detail in Section 4.3. Finally, by using the prior knowledge and the compression sensing reconstruction algorithm, we can obtain the estimated spectral support $\widehat{\Lambda}$. It is worth noting that, in the reconstruction algorithm, we can adaptively adjust the decision threshold by using the estimated noise intensity I_{Noise}.

4. Preprocessing of ABRMB Scheme

4.1. Estimation of Noise Intensity. Singular value decomposition is one of the most basic and important tools in modern numerical analysis. The SVD is performed on the measured values Y with noise using (6). $Y \in \mathbb{R}^{m\times r}$, $m \geq i > 0$, $r \geq j > 0$, and $m < r$, where r is sampling length.

$$Y = U\Sigma V^T. \tag{6}$$

In (6), unitary matrix $U \in \mathbb{R}^{m\times m}$ is the left singular vector of Y, and unitary matrix $V \in \mathbb{R}^{r\times r}$ is right singular vector. $\Sigma \in \mathbb{R}^{m\times r}$ is a diagonal matrix, and the main diagonal element is the singular value with a descending order. Since $\text{rank}(Y) = m$, (6) can be simplified as

$$Y = U_{\downarrow m} \begin{pmatrix} \Sigma_K & 0 \\ 0 & \Sigma_{m-K} \end{pmatrix} V_{\downarrow m}{}^T. \tag{7}$$

Let $\Sigma_K = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_K)$ and $\Sigma_{m-K} = \text{diag}(\sigma_{K+1}, \sigma_{K+3}, \ldots, \sigma_m)$. σ_i is ith singular value of Y, and $m \geq i \geq 1$. After SVD, an important characteristic is that most of the energy of signal is concentrated in the first K singular values, while the energy of the noise is distributed in all of the singular values; but noise can be reflected in the tail singular value. Assume that the singular value of Y can be decomposed into the singular value Σ_s of the original signal and the singular value Σ_n of the noise, and $\Sigma_s = \text{diag}(\sigma_{s1}, \sigma_{s2}, \ldots, \sigma_{sm})$, $\Sigma_n = \text{diag}(\sigma_{n1}, \sigma_{n2}, \ldots, \sigma_{nm})$. Figure 3 describes the contribution of Σ_s and Σ_n to Σ under different SNR, which are represented by R_s and R_n respectively. The definitions are as follows:

$$\begin{aligned} R_s &= \frac{\Sigma_s}{\Sigma} = \bigcup_{i=1}^{m} \left(\frac{\sigma_{si}}{\sigma_i} \right), \\ R_n &= \frac{\Sigma_n}{\Sigma} = \bigcup_{i=1}^{m} \left(\frac{\sigma_{ni}}{\sigma_i} \right). \end{aligned} \tag{8}$$

The contribution of the original signal to the tail of the singular values is small. The energy of the signal is concentrated in the first K singular values, and the tail of the singular values is mainly determined by noise. Figure 3 shows that the best data source for noise intensity estimation is bottom 50% of the singular values. Because the available singular values are also reduced with the decrease of the sampling channels, in order to ensure the accuracy of estimation, the bottom 30% of the singular values is adopted to estimate the noise intensity in this paper.

Figure 4 illustrates the variation of the singular values of Y with different noise intensity and different sampling channel number. As shown in Figure 4, the singular value is larger when the SNR is lower or the sampling channels number is more. It should be noted that since the initial part of the singular value is mainly determined by the energy of the signal, the initial part of the singular value has little change versus different SNR and channel numbers. Yet the tail of the singular values changes a lot under different SNR because it is influenced significantly by noise intensity as the contribution of real signal is much less for those values compared with noise. This fact is utilized in this paper to estimate the noise intensity.

Next, the distribution of the noise singular values is investigated. Since the noise is the uniformly distributed white Gaussian noise (WGN), it can be seen in Figure 5, the singular values of noise are largely distributed on a straight line under the circumstances with different SNR and different sampling channel number. Using this fact, the singular values of tail 30% of Y are used as singular values of noise to perform linear fitting. The fitting line is used to estimate the noise singular values, and the noise intensity is estimated by (9). The fitting result is shown in Figure 6. Since there is little

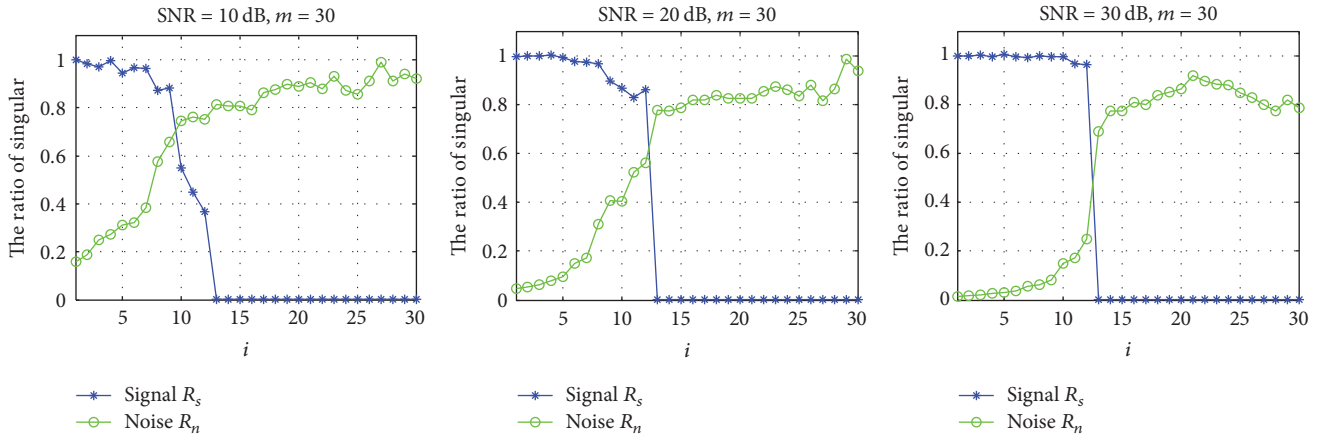


Figure 3: The contribution of the singular values of the signal and noise to total singular values.

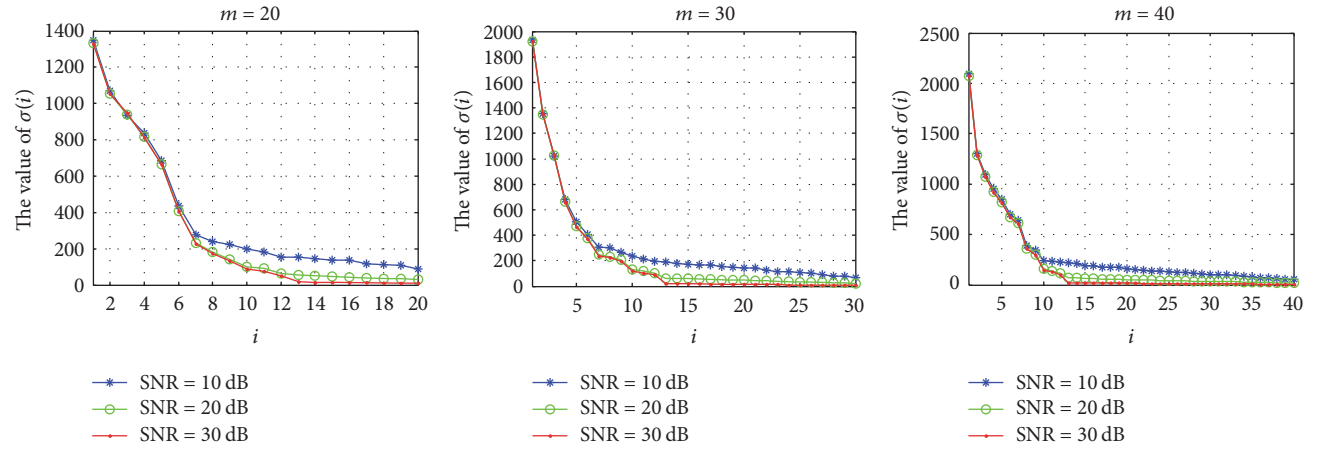


Figure 4: The comparison of singular value of Y under different SNR.

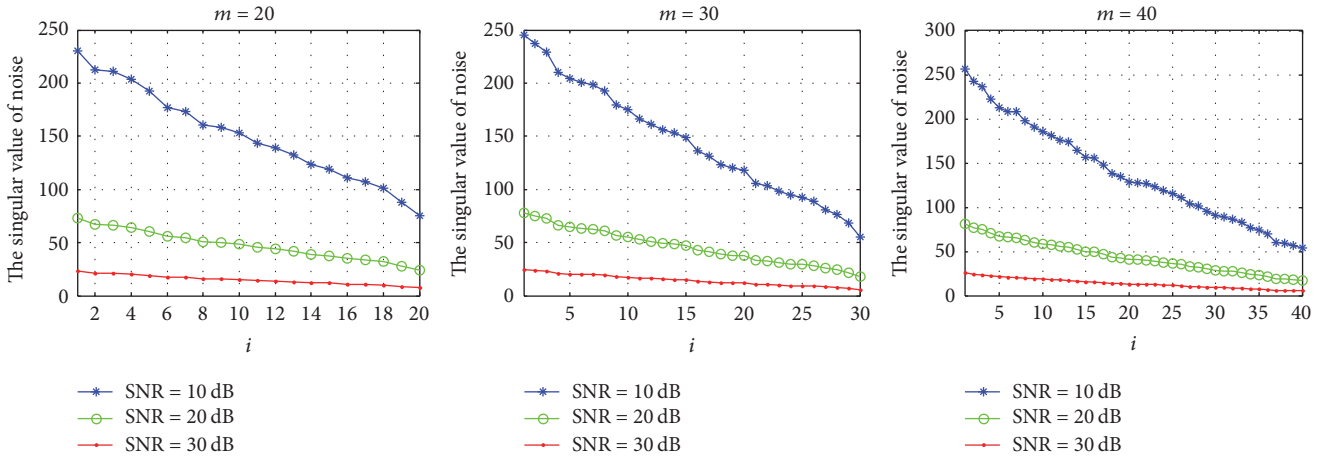


Figure 5: The distribution of noise singular values.

energy of real signal in the tail of the singular values, there is a certain deviation between the fitted singular values and original singular values of noise. However, under the low SNR, the deviation will be diluted due to the relatively large noise intensity. In addition, in order to reduce the influence of the deviation, this paper uses the optimal neighborhood selection strategy in Section 3.3 to carry out fault tolerance processing.

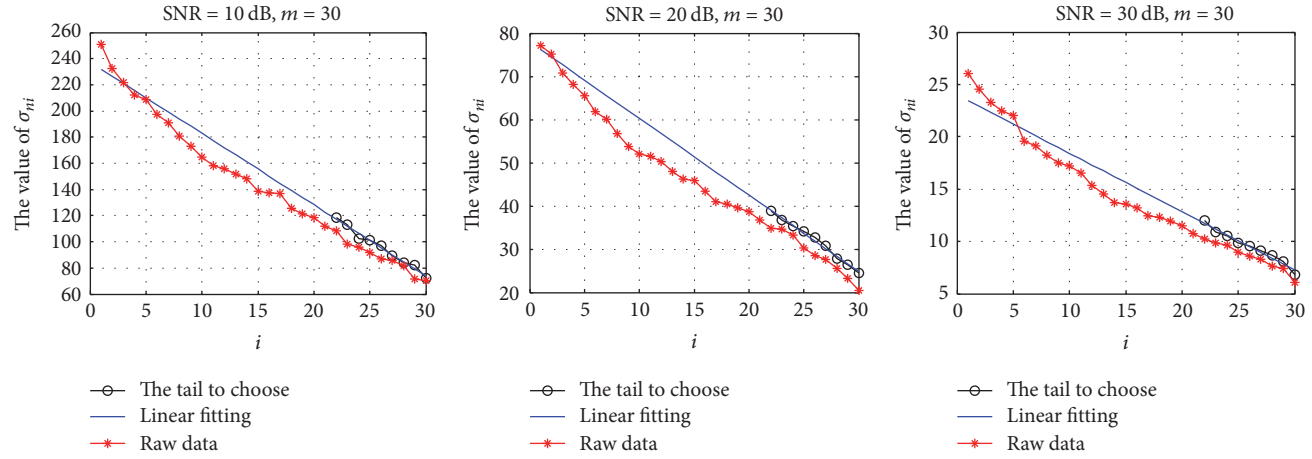


Figure 6: The linear fitting of the noise singular values.

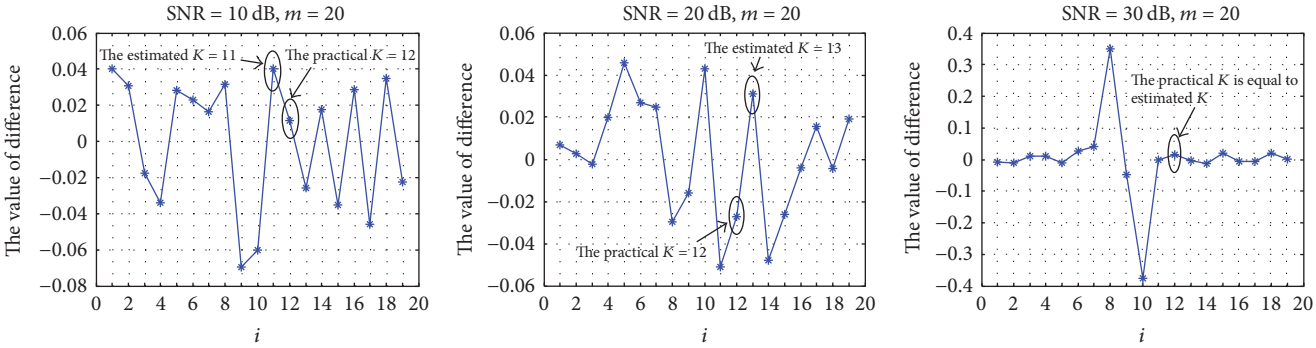


Figure 7: The sparsity order estimation under different SNR.

Suppose Nosie $\in \mathbb{R}^{m\times r}$ is a uniformly distributed WGN, then the intensity I_{Nosie} of the noise can be calculated according to the formula

$$\begin{aligned} \text{Nosie} &= US_nV^T \\ I_{\text{Noise}} &= \sqrt{\sum_{i=1}^{m} s_n{}^2(i)}, \end{aligned} \tag{9}$$

where $S_n = \{s_n(1), s_n(2), \ldots, s_n(i), \ldots s_n(m)\}$; $s_n(i)$ denotes the ith noise singular value.

4.2. Estimation of Sparsity Order. According to the previous analysis, we can get the estimated noise singular values $\widehat{\Sigma}_n = \text{diag}(\widehat{\sigma}_{n1}, \widehat{\sigma}_{n2}, \ldots, \widehat{\sigma}_{nm})$. The contribution of noise singular values on the singular values of Y is calculated by

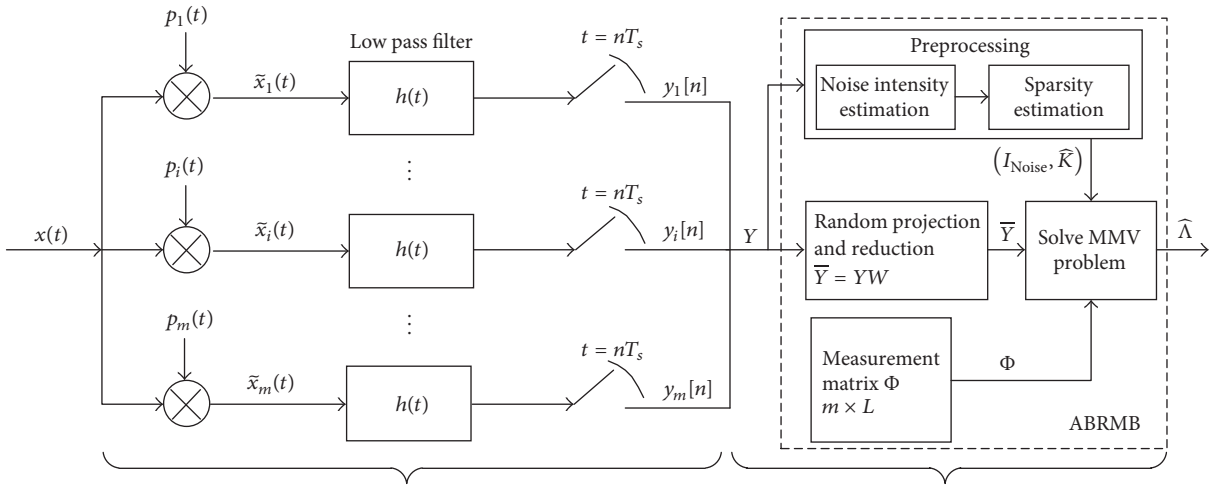
$$\widehat{R}_n = \frac{\widehat{\Sigma}_{n\cdot}}{\Sigma} = \bigcup_{i=1}^{m}\left(\frac{\widehat{\sigma}_{ni}}{\sigma_i}\right). \tag{10}$$

If the noise intensity is strong, G_{R_n} is first obtained by performing gradient operation on $\widehat{R}_n$, and then D_{R_n} is obtained by performing difference operation on G_{R_n}. The results of the operation are in an ascending order. The position of the minimum value in D_{R_n} plus 1 is the estimated sparsity order $\widehat{K}$. If the sampling channel number is close to the theoretical lower limit, $\widehat{K}$ need to add one adjustable parameter e, and empirical value e is 1. The calculation methods of G_{R_n} and D_{R_n} are shown in (11). Figure 7 is a sketch of estimation of sparsity order under different SNR.

$$\begin{aligned} G_{R_n} &= \text{abs}\left(\text{gradient}\left(\widehat{R}_n, 1\right)\right), \\ D_{R_n} &= \text{diff}\left(G_{R_n}\right). \end{aligned} \tag{11}$$

When noise is low, the signal energy is dominant. The sparsity order can be estimated directly by using the singular values of Y. First, all singular values are shifted to the left one time, and the last empty position is filled by σ_m, represented as $\Sigma_a = \text{diag}(\sigma_2, \sigma_3, \ldots, \sigma_m, \sigma_m)$. Then, R is calculated by (12) and is listed in a descending order. The index of maximum value in R is the estimated sparsity order $\widehat{K}$. Although $\widehat{K}$ has some deviations, the deviation has little impact on the success rate of reconstruction due to the joint sparse reconstruction and fault tolerance mechanism.

$$R = \frac{\Sigma_{a\cdot}}{\Sigma} = \bigcup_{i=1}^{m}\left(\frac{\sigma_{ai}}{\sigma_i}\right). \tag{12}$$



(a) MWC sub-Nyquist sampling (b) Support reconstruction

Figure 8: The support reconstruction scheme based on ABRMB.

4.3. Optimal Neighborhood Selection Strategy. From the signal model showed in Figure 1(a), we know that the energy of subband can only locate on two adjacent narrow channels. If support Λ_{i1} is obtained, support Λ_{i2} can only be its neighborhood channel; that is, $\Lambda_{i2} \in \text{nei}(\Lambda_{i1})$. However, there are usually two neighborhood channels of Λ_{i1}. Obviously, the channel with the largest norm 2 value is chosen to be the other support channel, as given in

$$\Lambda_{i2} = \text{pos}\left(\max\left(\left\|Z^{\text{nei}(\Lambda_{i1})\rightarrow}\right\|_2\right)\right), \tag{13}$$

where the function of pos() is to obtain the optimal neighborhood support of Λ_{i1}. $\|Z^{\text{nei}(\Lambda_{i1})\rightarrow}\|_2$ denotes the signal energy on two neighborhood channels of Λ_{i1}. Obviously, this strategy is very helpful to obtain the approximate spectral support.

5. Proposed ABRMB Optimization Scheme

5.1. Structure of ABRMB Scheme. It was pointed out in [18, 26] that random projection can be used for joint reconstruction to improve reconstruction performance. In particular, it can be known from [18] that the sensing performance tends to be stable if the number of preserved measured vectors is equal to N after the operation of projection. The ABRMB also inherits this idea, and it is combined with the pretreatment methods mentioned in Section 3 to improve the integrated system performance. The structure of ABRMB is shown in Figure 8.

5.2. Pseudocode of ABRMB Scheme. Pseudocode 1 provides the pseudocode to describe the ABRMB scheme. W represents a uniformly distributed random matrix whose values are continuously selected from $[-1, 1]$.

As can be seen in Pseudocode 1, the ABRMB allows a certain degree approximation deviation for spectral support with the aid of estimated noise intensity and sparsity order. In addition, if the loop iteration is ended, the optimal spectral support is still not found; the ABRMB can select the spectral support, with minimum potential, as the optimal support set from all of the stored spectral support. If there are more than one minimum potential, then choose the one with smallest residual as the best spectral support. Because of the guarantee of fault tolerance mechanism, the experiment shows that this method is very effective.

The solver of ABRMB scheme is described in Pseudocode 2.

5.3. The Convergence of ABRMB Scheme

Theorem 1. *Assume that $x(t)$ is multiband signal described as in Figure 1(a); we use the MWC structure for sample signal as shown in Figure 8. If the following conditions are established, for any $f \in [-f_s/2, f_s/2]$, $z(f)$ is the only N-sparse solution of (4).*

1. $f_s \geq f_p \geq B$, *and* $f_s/f_p < (M_{\min} + 1)/2$.
2. $m \geq 2N$.
3. *The number M of ± 1 symbols in a periodic sequence $p_i(t)$ must satisfy* $M \geq M_{\min} = \lceil f_{\text{nyq}}/f_p \rceil$. *If* $f_s = f_p$, *then* $M_{\min} = L$.
4. *Any $2N$ column of Φ is linearly independent.*

The proof of Theorem 1 is shown in [10].

Next, we prove that the spectral support of the original MMV problem $\Lambda = \text{supp}(Z)$ can be obtained by solving the new MMV problem $\overline{Y} = \Phi Z$, where $\overline{Y}$ is the result after projection.

Lemma 2. *Suppose $a \in R^L$ is a known nonzero vector, and $u \in R^L$ is a random set of vectors obtained from a continuous probability distribution. Then, the probability that the event "a on u is not 0" is 1.*

Input: Observation matrix Φ, compressed sampling signal Y, reconstruction error ε, tuning parameter s of the potential of spectral support, maximum iteration MaxIters.
Initialization: iter = 1.
Pretreatment:
(1) Noise intensity estimation: $\widehat{I}_{\text{Nosie}} = \text{norm}(\widehat{\Sigma}_n)$.
(2) Sparsity order estimation: pos = sort $(D_{R_n}, \text{'ascend'})$, $\widehat{K} = \text{pos}(1) + 1 + e$.
(3) The number of sub-bands: $\widehat{N} = \text{ceil}(\widehat{K}/2)$.
Iteration:
While iter ≤ MaxIters
(1) Random projection and dimension reduction: $W = \text{random}(-1, 1, r, \widehat{N})$, $\overline{Y} = YW$.
(2) Solver solution: $\overline{Y} = \Phi Z$, $[\widehat{\Lambda}, \widehat{Z}] = \text{ABRMB\_Solver}(\overline{Y}, \Phi, \widehat{K}, \widehat{I}_{\text{Noise}})$.
(3) Storage of spectral support: $\text{supp}(\text{iter}) = \widehat{\Lambda}$.
(4) Obtain the estimation and calculate the residual: $\overline{Y}_e = \Phi\widehat{Z}$,
$\text{residual} = (\text{norm}(\overline{Y} - \overline{Y}_e)/\text{norm}(\overline{Y}_e)/(1 + \widehat{I}_{\text{Nosie}}))$.
(5) Condition test: If $(|\widehat{\Lambda}| \le \widehat{K} + s)\&\&\text{residual} \le \varepsilon$
Break;
End
(6) iter = iter + 1.
End
Until the end of the iteration, condition (5) is not satisfied, then make the best selection:
If iter > MaxIters
$\widehat{\Lambda} = \text{find}(\min(\text{len}(|\text{supp}_{j\downarrow}|)))$.
End
Output: return support $\widehat{\Lambda}$

Pseudocode 1: Pseudocode of ABRMB scheme.

Input: $\overline{Y}, \Phi, \widehat{K}, \widehat{I}_{\text{Noise}}$.
Initialization: $\widehat{\Lambda}_0 = \emptyset$, RowNorm = $\emptyset$.
(1) OMPMMV is used to perform joint reconstruction, solve $\overline{Y} = \Phi Z$. The preliminary solution $\dot{Z}$ is obtained.
(2) $\text{RowNorm} = \text{norm}(\dot{Z}^{i\rightarrow})$.
(3) Initial spectral support: $\dot{\Lambda} = \max(\text{RowNorm}, \widehat{K})$.
(4) Optimal neighborhood selection: $\widehat{\Lambda}_0 = \dot{\Lambda} \cup \text{pos}(\max(\|\dot{Z}^{\text{nei}(\Lambda_0)\rightarrow}\|_2))$.
(5) Obtain a better solution, and calculate norm l_2: $\ddot{Z} = \Phi^{\dagger}_{\downarrow\widehat{\Lambda}_0}\overline{Y}$, $\text{RowNorm} = \text{norm}(\ddot{Z}^{i\rightarrow})$.
(6) Obtain optimal spectral support and optimal solution: $\ddot{\Lambda} = \text{find}(\text{RowNorm} >= \widehat{I}_{\text{Noise}})$,
$\dddot{Z} = \Phi^{\dagger}_{\downarrow\ddot{\Lambda}}\overline{Y}$
Output: $\ddot{\Lambda}, \dddot{Z}$.

Pseudocode 2: Pseudocode of ABRMB_Solver.

Theorem 3. *We know $\sigma(\Phi) \ge 2K$, where $\sigma(\Phi)$ is the Kruskal rank of Φ, $\widehat{Z}$ is the only K-sparse solution of $Y = \Phi Z$, and $W = \{w_1, w_2, \ldots, w_J\}$ $(1 \le J \le r)$ is a random unit vector set obtained from a continuous probability distribution, whose length is r. Suppose $\overline{Y} = YW$, $\overline{Z} = \widehat{Z}W$, considering a new MMV problem $\overline{Y} = \Phi Z$ after dimension reduction, the probability of* $\text{supp}(\widehat{Z}) = \text{supp}(\overline{Z})$ *is 1. Meanwhile, $\overline{Z}$ is the only K-sparse solution.*

The proof of Lemma 2 and Theorem 3 can be found in [18].

Thus, Theorem 1, Lemma 2, and Theorem 3 can guarantee the convergence of the proposed scheme.

6. Numerical Results and Performance Analysis

6.1. Performance Evaluation Indices

6.1.1. The Required Minimum Number of Sampling Channels and Minimum Sampling Rate. When the required conditions are met, it is pointed out in [10] that the spectral support of the signal can be reconstructed as long as the number of sampling channels is satisfied by $m \ge 2N$ in theory. However, the existing algorithms are difficultly reach the theoretical lower limit and can only approximate the theoretical value by continuously improving the algorithms.

In addition, MWC can achieve sub-Nyquist sampling. Without affecting the success rate of reconstruction, the subsequent processing of the system is more favorable with higher degree of compression. The total sampling rate of MWC is $f_\Sigma = mf_s$. The theoretical minimum of sampling rate for multiband signal, that is, the Landau rate [15], is defined as

$$M(P_N) = 2\sum_{i=1}^{N/2}(b_i - a_i), \tag{14}$$

where $M(P_N)$ is Landau rate; that is, it is the sum of all subband frequency widths, P_N denotes the union of N effective frequency components in multiband signal, and $(b_i - a_i)$ is frequency width of ith subband.

Since the number of the channels m and total sampling rate f_Σ are directly relevant, obviously, the cost of the system and the corresponding sampling rate are lower for smaller m.

6.1.2. The Maximum Number of Subbands That Can Be Reconstructed in Signal. From [16], the upper bound of the reconstruction capability of the MMV problem is

$$I(Z) \le \frac{(\sigma(\Phi) + \mathrm{Rank}(Y) - 1)}{2}. \tag{15}$$

In (15), $I(Z)$ is the joint sparsity order of Z, $\sigma(\Phi)$ is the Kruskal rank of Φ, and $\mathrm{Rank}(Y)$ is the rank of Y. As can be learned from (15), the smaller the rank Y, the lower the sparsity order of the signal that can be reconstructed, which is also the reason that the upper bound of the sparsity order in [15] is particularly small. The number of subbands is $N = I(Z)/2$. Therefore, the performance of the system is better when the number of subbands that can be reconstructed is larger.

6.1.3. The Approximate Error of the Support Potential. For the original signal, the potential of spectral support is $|\Lambda|$, that is, the length of spectral support. Obviously, $|\Lambda| = \|\Lambda\|_0$. If $|\widehat{\Lambda}| = |\Lambda|$, then the approximate error is 0. If $|\widehat{\Lambda}| > |\Lambda|$, then the approximate error existed. In this paper, the approximate error $E_{\widehat{\Lambda}}$ and the upper bound of the error E_{upper} are defined as

$$E_{\widehat{\Lambda}} = \frac{\left(\left|\widehat{\Lambda}\right| - |\Lambda|\right)}{L} \quad \text{s.t. } \left|\widehat{\Lambda}\right| > |\Lambda| \tag{16}$$

$$E_{\mathrm{upper}} = \frac{|\Lambda|_{\max}}{L}. \tag{17}$$

In (16), $|\widehat{\Lambda}|$ is the potential of estimated support, $(|\widehat{\Lambda}| - |\Lambda|)$ represents the difference between the estimated support potential and the actual support potential, and "s.t." is the abbreviation of "subject to". In (17), $|\Lambda|_{\max}$ is the potential of maximum spectral support for multiband sparse signals. L is the number of narrow bands, and N is the number of support bands of the original signal. Apparently, smaller approximate error is better. Since $L \gg |\Lambda|$, in the application of wideband spectrum sensing in cognitive radio, as long as $E_{\widehat{\Lambda}} \le E_{\mathrm{upper}}$, the impact on the secondary users can be negligible.

6.1.4. The Successful Probability of the Reconstruction. In the analysis of recovered spectral support, we refer to the successful recovery criteria in [10]; that is, when the estimated support $\widehat{\Lambda}$ and the actual support Λ satisfy (18), where $\widehat{\Lambda} \supseteq \Lambda$, and $\Phi_{\downarrow\widehat{\Lambda}}$ is with full column rank, it is considered as a successful reconstruction. In the experiment of this paper, reconstruction is performed 500 times in the same simulation environment. If the success rate of reconstruction is more than 90%, it is considered a high probability reconstruction.

$$\text{success} \quad \text{s.t. } \left(E_{\widehat{\Lambda}} \le E_{\mathrm{upper}} \&\& \widehat{\Lambda} \supseteq \Lambda \&\& \mathrm{Rank}\left(\Phi_{\downarrow\widehat{\Lambda}}\right) = \left\|\widehat{\Lambda}\right\|_0\right). \tag{18}$$

6.2. Experimental Analysis. In order to verify the effectiveness of the proposed scheme, simulation experiments are carried out to analyze the system performance metrics, including minimum number of sampling channels, the success rate of reconstruction, and the maximum number of subbands that can be reconstructed. Then, these metrics are used to compare with ReMBo and PRMB algorithms. Meanwhile, the approximate error of support potential is also analyzed.

The sparse multiband analog signal with noise is generated from sinc signal by

$$x(t) = \sum_{i=1}^{N/2}\sqrt{E_i B_i}\,\mathrm{sinc}\left(B_i(t - \tau_i)\right)\cos\left(2\pi f_i(t - \tau_i)\right) + n(t). \tag{19}$$

In (19), E_i, B_i, f_i, and τ_i represent the energy coefficient, bandwidth, carrier frequency, and time offset of the ith band, respectively. N is the number of subbands in the signal. $n(t)$ is white Gaussian noise. The following procedure is repeated 500 times to calculate the probability of success.

1. Generate the mixing signal $p_i(t)$ randomly.
2. Generate the carrier frequency f_i uniformly and randomly in $[-f_{\mathrm{nyq}}/2, f_{\mathrm{nyq}}/2]$.
3. Generate new sinc signal according to f_i.
4. Estimate the spectral support using ReMBo, RPMB, and ABRMB, respectively, and determine whether it is successfully recovered.

In the simulation, the settings of parameters of the signal are $N = 6$ (3 balanced pairs); $E_i \in \{1, 2, 3\}$, $B_i \in \{50, 50, 50\}$ MHz, $\tau_i \in \{0.4, 0.7, 0.2\}$ μs, carrier frequency f_i randomly distributed in the interval $[-f_{\mathrm{nyq}}/2, f_{\mathrm{nyq}}/2]$, and $f_{\mathrm{nyq}} = 10$ GHz. The parameters of MWC sampling are $L_0 = 97$, $L = 2L_0 + 1 = 195$, $f_s = f_p = f_{\mathrm{nyq}}/L = 51.28$ MHz, number of channels m steps using 1 as interval in the range $[15, 40]$, and MaxIters = 20. The noise is WGN, and SNR = $\{10, 20, 30\}$ dB. The other parameters are $\varepsilon = 0.001$, $s = 2$.

Firstly, under the same conditions, the changes of the success rate of reconstruction with channel number using ABRMB, RPMB, and ReMBo are studied under different SNR. As can be seen in Figure 9, when SNR equals 10, 20, and 30, respectively, the success rate of the reconstruction of

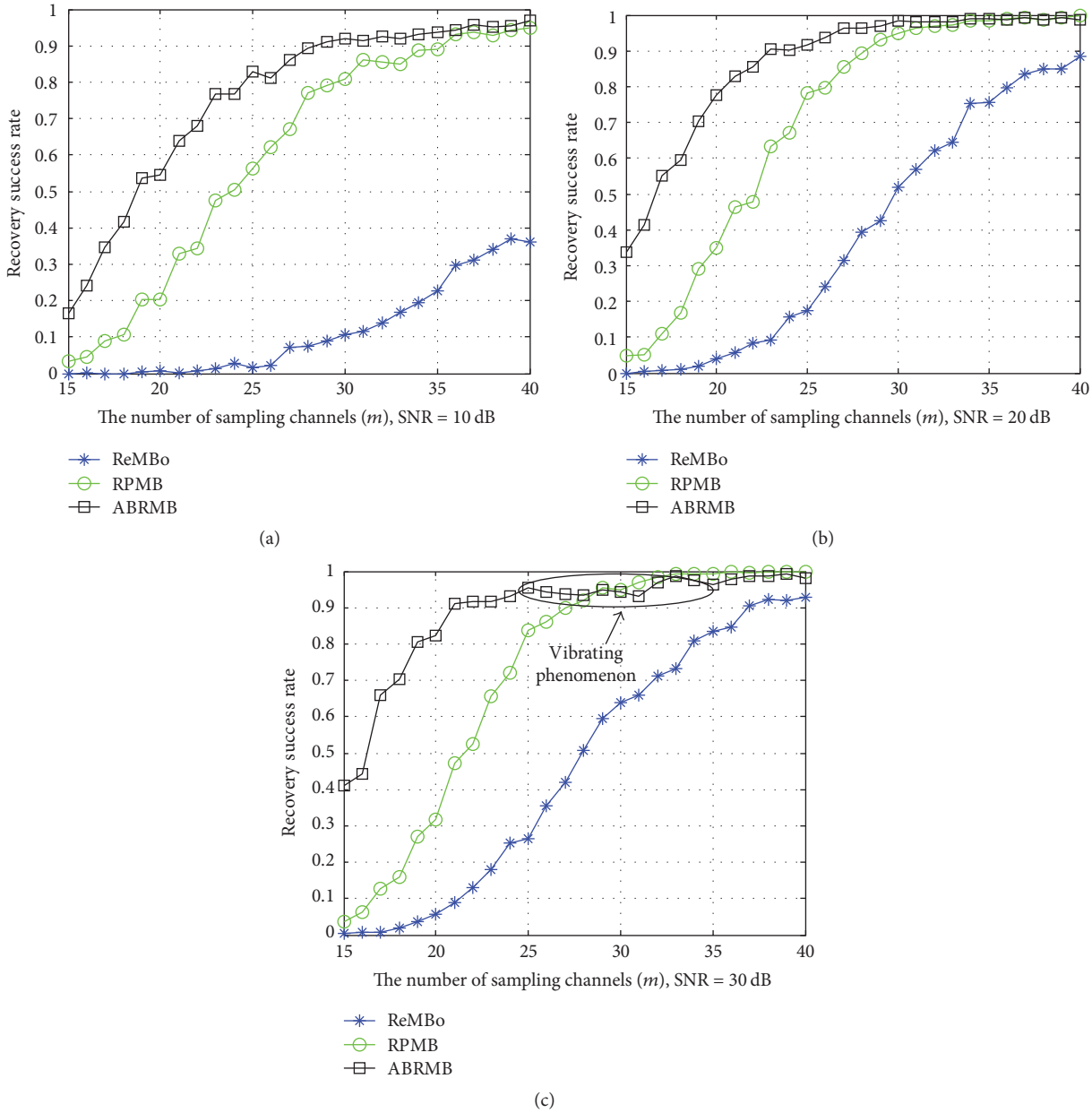


Figure 9: The comparison of recovery success rate of support under different SNR.

spectral support using ABRMB is better than using ReMBo or RPMB. Particularly, when $m = 20$ and SNR $= \{10, 20, 30\}$ dB, the improvements of the success probability of ABRMB compared with RPMB and ReMBo are shown in Table 1. Therefore, ABRMB scheme is more effective and has better adaptability in the case when SNR is low or fluctuated due to the reasonable estimation of the noise intensity and the sparsity order and the optimal neighborhood selection strategy.

Secondly, in the same circumstances, the needed minimum channel number and the minimum sampling rate of high probability reconstruction are studied. As shown in Figure 9 and Table 2, under different SNR, the number of hardware channels and the sampling rate of ABRMB for the high probability reconstruction are smaller than those of the ReMBo and RPMB. Thus, our proposed scheme can use fewer hardware channels and lower sampling rate to achieve the high success rate of reconstruction. Obviously, the cost of

Table 1: Comparison of maximum promoting rate of the reconstruction success probability.

Comparison	$m = 20$		
	SNR = 10 dB	SNR = 20 dB	SNR = 30 dB
ABRMB versus RPMB	34.2%↑	42.8%↑	50.8%↑
ABRMB versus ReMBo	53.6%↑	73.6%↑	76.6%↑

Table 2: The comparison of the minimum channel and minimum sampling rate needed.

Algorithms	SNR = 10 dB		SNR = 20 dB		SNR = 30 dB	
	$m_{\min}$	$f_{\sum \min}$/MHz	$m_{\min}$	$f_{\sum \min}$/MHz	$m_{\min}$	$f_{\sum \min}$/MHz
ABRMB	29	1487.12	23	1179.44	21	1076.88
ReMBo	–	–	40	2051.2	37	1897.36
RPMB	36	1846.08	28	1435.84	27	1384.56

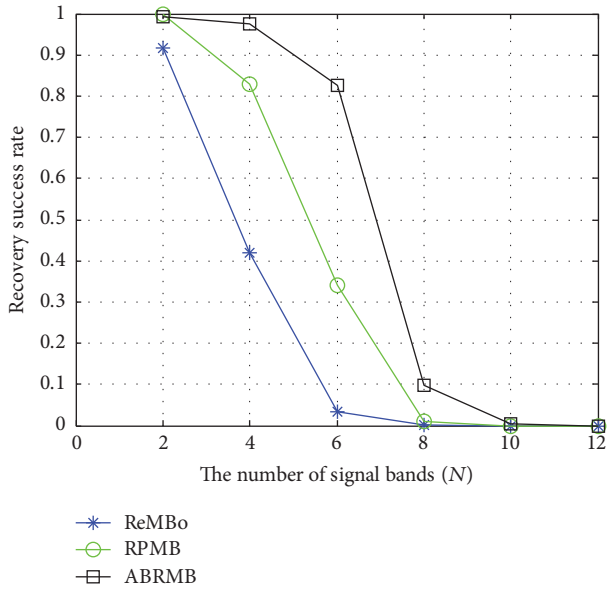


Figure 10: The effects of subbands number on the support recovery.

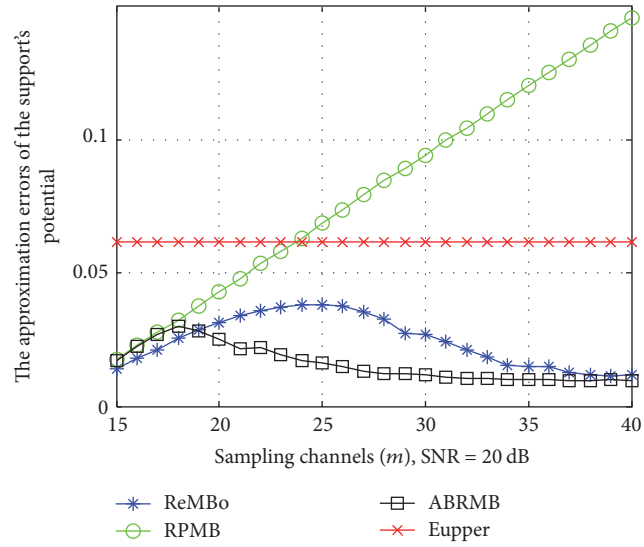


Figure 11: The comparison of the approximation error of the support potential.

the system can be saved as the number of hardware channels needed by the scheme which can be reduced.

Next, under the same conditions, the changes of success rate of reconstruction with the number of the bands are studied using these three algorithms. The number of the subbands N is chosen in the range $[2, 12]$ with 2 as its advancing step. The parameters are SNR = 20 dB, $m = 20$, $E_i \in \{1, 2, 3, 4, 5, 6\}$, and $\tau_i \in \{0.4, 0.7, 0.2, 0.9, 1.2, 1.5\}$ μs. As shown in Figure 10, since the number of the bands is directly related to the sparsity order of the signal, with the increase of N, the signal is not sparse enough and the success rate of reconstruction for three algorithms is reduced. Nevertheless, the reconstruction performance of ABRMB is obviously better than that of RPMB and ReMBo for $N = 4$, 6, and 8.

At last, we compare the average approximation error on the potential of estimated spectral support. We set $N = 6$, SNR = 20 dB, and m is in the range $[15, 40]$ with 1 as its step. As can be seen in Figure 11, since the potential of the support reconstructed by RPMB is related to the number of channels, when $m = 24$, the average approximation error of the spectral support of RPMB has already reached the upper limit. From Figure 11, the average approximation error of ABRMB is the smallest in these three algorithms, and the potential of the spectral support reconstructed by ABRMB is the closest to the actual number of frequency channels. In this way, the ABRMB can provide more spectrum access opportunities for the secondary users in cognitive radio networks.

From the analysis of above four performance metrics, we can see that ABRMB scheme can achieve low-speed sampling by utilizing the sparse characteristics of wideband signals in frequency domain. The sampling rate can be reduced to 14.9% of the Nyquist sampling rate when SNR = 10 dB. Meanwhile, for noise uncertainty, the detection threshold can be adaptively changed as the noise power changes in ABRMB scheme. Based on the neighborhood selection strategy and sparsity estimation, ABRMB can obtain higher success reconstruction probability. Then, the projection reduction operation can reduce the computational complexity, and it does not affect the convergence of the scheme. Finally, ABRMB scheme can find better trade-off between noise interference and sensing accuracy.

7. Conclusions

By using SVD theory, this paper proposes a self-adaptive and blind wideband spectrum sensing MWC scheme which leads to a flexible and high performance solver. The SVD is performed before signal reconstruction in the preprocessing block. The noise intensity and sparsity order of the signal are estimated in this block, and then the subbands number of the signal can be obtained. ABRMB scheme can use the estimated noise intensity and sparsity order to process the multiband

signals adaptively, and it improves the reconstruction performance by using optimal neighborhood selection. In practical sensing environment, especially for low or fluctuated SNR, the reconstruction of the spectral support can be performed more accurately by using the ABRMB. The simulation results demonstrate that, compared with ReMBo algorithm and RPMB algorithm, the reconstruction performance has been improved by 34.2–76.6%; the minimum number of hardware channels and minimum sampling rate needed by high probability reconstruction have been decreased by 17.9–22.2%. When we perform reconstruction of spectral support on the signal with different number of subbands, the performance of ABRMB is better than the other two algorithms. In the study of the approximation error of the support potential by using the three algorithms, the approximation error of ABRMB is the smallest.

Notations

Λ: The actual spectral support of signal
$\dot{\Lambda}$: The obtained initial support in ABRMB_Solver
$\ddot{\Lambda}$: The optimal solution in ABRMB_Solver
$\widehat{\Lambda}$: The obtained spectral support with ABRMB
$|\Lambda|$: The potential of spectral support
N: The number of subbands in multiband signals
B_i: Bandwidth of ith subband
f_{nyq}: Nyquist rate of $x(t)$
$p_i(t)$: Periodic mixing signal
Y: The sub-Nyquist sampling signal with MWC
$\overline{Y}$: The sampling matrix after projection
K: Sparsity order of signal
$\widehat{K}$: The estimated sparsity order
$\|\widetilde{Z}^{i\rightarrow}\|_2$: The norm of each row vector for $\widetilde{Z}$
Φ: $m \times L$ measurement matrix or observation matrix
$\Phi^{\dagger}{}_{\downarrow\ddot{\Lambda}}$: Extracting column vectors from Φ according to $\ddot{\Lambda}$ and then conducting pseudoinverse operation
ε: The decision threshold
I_{Nosie}: The noise intensity
E_i: The energy coefficient of ith subband
f_i: Carrier frequency
τ_i: The time offset of ith subband
ceil: The function of round toward positive infinity
L: Spectrum slice number
f_p: Spectral slice width, $f_p = f_{\text{nyq}}/L$
f_s: Sampling rate at each channel, $f_s = qf_p$, with odd q.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61561017), the Hainan Province Natural Science Foundation of China (617033, 20166232), the Open Sub-Project of State Key Laboratory of Marine Resource Utilization in South China Sea (2016013B), Oriented Project of State Key Laboratory of Marine Resource Utilization in South China Sea (DX2017012), and the Major Science and Technology Project of Hainan Province (no. ZDKJ2016015).

References

[1] H. Sun, A. Nallanathan, C. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Communications Magazine*, vol. 20, no. 2, pp. 74–81, 2013.

[2] A. Furtado, L. Irio, R. Oliveira, L. Bernardo, and R. Dinis, "Spectrum sensing performance in cognitive radio networks with multiple primary users," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1564–1574, 2016.

[3] A. A. Alkheir and M. Ibnkahla, "A selective decision–fusion rule for cooperative spectrum sensing using energy detection," *Wireless Communications and Mobile Computing*, vol. 16, no. 12, pp. 1603–1611, 2016.

[4] S. K. Sharma, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Application of compressive sensing in cognitive radio communications: a survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 1838–1860, 2016.

[5] M. Yang, Y. Li, X. Liu, and W. Tang, "Cyclostationary feature detection based spectrum sensing algorithm under complicated electromagnetic environment in cognitive radio networks," *China Communications*, vol. 12, no. 9, Article ID 7275257, pp. 35–44, 2015.

[6] K. Cichon, A. Kliks, and H. Bogucka, "Energy-efficient cooperative spectrum sensing: a survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 1861–1886, 2016.

[7] M. J. Saber and S. M. S. Sadough, "Multiband cooperative spectrum sensing for cognitive radio in the presence of malicious users," *IEEE Communications Letters*, vol. 20, no. 2, pp. 404–407, 2016.

[8] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. 1357–1360, April 2007.

[9] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[10] M. Mishali and Y. C. Eldar, "From theory to practice: sub-nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, 2010.

[11] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: analog to digital at sub-Nyquist rates," *IET Circuits, Devices and Systems*, vol. 5, no. 1, pp. 8–20, 2011.

[12] Y. C. Eldar, R. Levi, and A. Cohen, "Clutter removal in sub-Nyquist radar," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 177–181, 2015.

[13] M. A. Lexa, M. E. Davies, and J. S. Thompson, "Reconciling compressive sampling systems for spectrally sparse continuous-time signals," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 155–171, 2012.

[14] D. Cohen, A. Akiva, B. Avraham, S. Patterson, and Y. C. Eldar, "Distributed cooperative spectrum sensing from sub-Nyquist samples for Cognitive Radios," in *Proceedings of the 16th IEEE International Workshop on Signal Processing Advances in Wireless Communications, SPAWC '15*, pp. 336–340, IEEE, Stockholm, Sweden, July 2015.

[15] M. Mishali and Y. C. Eldar, "Reduce and boost: recovering arbitrary sets of jointly sparse vectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, part 1, pp. 4692–4702, 2008.

[16] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[17] H. J. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 25, no. 2, pp. 37–52, 1967.

[18] J. X. Gai, P. Fu, J. Y. Sun, H. J. Lin, and L. H. Wu, "A recovery algorithm of MWC sub-Nyquist sampling based on random projection method," *Acta Electronica Sinica*, vol. 42, no. 9, pp. 1686–1692, 2014.

[19] Z. Drmac, "Accurate computation of the product-induced singular value decomposition with applications," *SIAM Journal on Numerical Analysis*, vol. 35, no. 5, pp. 1969–1994, 1998.

[20] X. Yang, X. Tao, Q. Cui, and Y. J. Guo, "nterference-constrained adaptive simultaneous spectrum sensing and data transmission scheme for unslotted cognitive radio network," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 102, 2012.

[21] N. Hasan, H. Kim, W. Ejaz, and S. Lee, "Knapsack-based energy-efficient node selection scheme for cooperative spectrum sensing in cognitive radio sensor networks," *IET Communications*, vol. 6, no. 17, pp. 2998–3005, 2012.

[22] W. Ejaz, G. A. Shah, N. U. Hasan, and H. S. Kim, "Energy and throughput efficient cooperative spectrum sensing in cognitive radio sensor networks," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 7, pp. 1019–1030, 2015.

[23] Z. H. Hu, Y. Bai, Y. C. Zhao, and Y. R. Zhang, "Support recovery for multiband spectrum sensing based on modulated wideband converter with SwSOMP algorithm," in *Proceedings of the 1st EAI International Conference on 5G for Future Wireless Networks*, 2017.

[24] W. Sun, Z. Huang, F. Wang, and X. Wang, "Compressive wideband spectrum sensing based on single channel," *IEEE Electronics Letters*, vol. 51, no. 9, pp. 693–695, 2015.

[25] P. Yang, Z.-T. Huang, Z. Liu, and W.-L. Jiang, "Single-channel spectrum sensing technique based on sub-Nyquist sampling," in *Proceedings of the 2012 11th International Conference on Signal Processing, ICSP '12*, pp. 224–227, October 2012.

[26] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, 2005.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 7254181, 9 pages
https://doi.org/10.1155/2017/7254181

Research Article

ABS-SmartPriority: An Agent-Based Simulator of Strategies for Managing Self-Reported Priorities in Smart Cities

Iván García-Magariño[1,2] and Raquel Lacuesta[1,2]

[1]*Department of Computer Science and Engineering of Systems, University of Zaragoza, Escuela Universitaria Politécnica de Teruel, c/Atarazana 2, 44003 Teruel, Spain*
[2]*Instituto de Investigación Sanitaria Aragón, University of Zaragoza, Zaragoza, Spain*

Correspondence should be addressed to Iván García-Magariño; ivangmg@unizar.es

Received 29 August 2017; Accepted 25 October 2017; Published 10 December 2017

Academic Editor: Syed Hassan Ahmed

Smart cities still need the appropriate tools for allowing researchers to contribute in this growing field that will change the comfort and qualities of live of citizens. Smart cities can provide services such as informing of the less overcrowded tourist routes, path-finding for avoiding traffic jams, search services for parking, collecting tree branches from streets, and the finding of the nearest available public electric bicycles or cars. The levels of urgencies differ from some situations to others. Examples of urgent matters are the fast transportation critical patients and the removal of obstacles in main roads. Since smart cities are meant to manage huge amounts of requests, priorities should be automated and managed in a flexible way for new scenarios. Smart cities may need citizens to report the service priorities. However, citizens may have different criteria and could abuse the highest priority levels hindering the service performance for really urgent matters. A novel agent-based simulation open-source framework is proposed for testing different policies for normalizing and controlling self-reported priorities, with its simulator called ABS-SmartPriority. This approach is illustrated by simulating two different policies, in which the smart policy outperformed the one used as the control mechanism.

1. Introduction

One of the main goals of smart cities is to improve the quality of service (QoS) of citizens [1]. In smart cities, information communication technologies (ICT) allow citizens to use more efficiently the existing resources. In particular, network is the main necessity for the communication of the devices with Internet of Things (IoT) and the different kind of sensors [2]. Among others, some examples of sensors are waste collection sensors [3], conductivity sensors for monitoring groundwater resources [4], and radio frequency identification (RFID) sensors for identifying users of smart cities [5].

The wireless sensor networks (WSNs) can be appropriate for supporting communications in smart cities thanks to their low-cost and their easily adaptive configuration to different city circumstances. However, maintaining QoS with WSNs is challenging due to the power constraints, radio frequency interferences, noise, and multipath finding. Yigit et al. [6] proposed a priority mechanism for addressing these challenges.

In the management of city services, a priority mechanism can make it possible that urgent matters are attended faster. For example, Anagnostopoulos et al. [7] used priorities for determining how to collect waste in smart cities, prioritizing the areas that can have potentially risky waste such as hospital areas. Another example is the way patients are commonly attended every day in a hospital. When a patient gets to the hospital for the emergency department, a healthcare professional almost immediately performs a really quick examination to determine the actual urgency of the matter to establish a priority level [8]. This saves lives, as life-critical matters are attended with the highest priority. This priority policy works properly thanks to the existence of a common individual that examines objectively all the cases and assigns priorities in a fair way.

Nevertheless, the smart cities show a steep exponential growth of data and communications with imprecise information [2]. This amount of information does not make it possible to have independent individuals to fairly assign the priorities for every service. Then, the assignment of priorities may need to be indicated by the citizens of the smart city or its mobile applications. However, the self-assigned priorities would require a great level of responsibility from behalf of citizens. The Prisoner's Dilemma reveals that the self-assigned priority in an anonymous way (in the sense that nobody notices the behavior of each citizen) may be useless. Most citizens could request the highest level of priority (priority zero) for all the requests of services regardless their actual urgency. Martínez-Cánovas et al. [9] showed that cooperation is promoted when users repeat the decisions knowing that these are tracked and can be rewarded (or penalized).

The current work belongs to the trend of works about providing testbeds for analyzing and simulating different scenarios of smart cities. For example, Sanchez et al. [10] provided a testbed for simulating and analyzing different strategies for managing the information from IoT devices in smart cities. Although their testbed was initially inspired by Santander city, their testbed is applicable to many situations of smart cities in general. In addition, the current work is also in the line of research about scheduling services in smart cities. For instance, Anagnostopoulos et al. [11] proposed a scheduling mechanism for the waste collection in smart cities. They analyzed and tested their approach using data from the St. Petersburg city, and they presented an Android app for evaluating the feasibility of their approach.

In this scenario, multiagent systems (MASs) allow developers to implement systems with several autonomous entities for testing behaviors of societies composed of several individuals. For example, MASs have been shown to be useful for studying (a) the coordination of citizens in urban crisis situations [12] and (b) the coordination a group of experts for reaching a consensus with the Delphi process [13]. More concretely, agent-based simulators (ABSs) are a specific kind of MASs that is aimed at analyzing different circumstances by simulations. For instance, the ABS of Hassan et al. [14] simulated the friendship and partner relationships taking proximity and similarity into account with fuzzy techniques. In addition, the ABS for Tourist Urban Routes (ABSTUR) [15] showed its utility for analyzing the emergent behavior of tourists visiting a city based on the recommendations of certain routes. Furthermore, Garcia-Sabater et al. [16] proposed an ABS to anticipate the demands of manufacturing networks.

In this context, we present a novel open-source ABS and its underlying framework, both of them called ABS-SmartPriority, as a solution to assess and compare different self-reported priority strategies in smart city societies. This approach allows researchers to explore new ways of fairly managing the resources and networks of smart cities considering self-reported priorities.

2. Materials and Methods

2.1. ABS-SmartPriority. This work uses the novel ABS called ABS-SmartPriority for simulating the management of resources of smart cities based on self-reported priorities. This can simulate priorities reported by citizens, city workers, or mobile applications on behalf of their users. This ABS is developed with the Process for developing Efficient ABS (PEABS) [17]. In order to ensure the reproducibility of the experiments, the source code of the underlying framework has been made available as supplementary material of the current article [18]. This open-source code may allow other researches to extend the existing framework or understand better the underlying implementation. The framework was programmed with the C# language and the Unity engine. We selected Unity because (a) it is a cross-platform engine that allows deploying applications in many platforms such as Windows, Mac, Android, and iOS and (b) it facilitates the programming of visual and interactive applications. From the two programming languages supported by Unity, we chose C# since its object-oriented nature facilitated the straightforward application of PEABS for implementing an efficient ABS. In addition, agents are normally implemented as objects in object-oriented programming in the literature [19].

The user can assign priorities in a range (in our case 0 to 2), in which zero represents the highest priority and two the lowest one. However, the framework could be easily adapted to represent different priority ranges. In the real world, the user would be a person that assigns this level of priority for any request of service or resource. In fact, the user could set default priorities for certain activities through their mobile applications for not being annoyed in selecting the priority in most cases. However, since this work is intended to simulate different behaviors, each person is impersonated by an agent, which will be referred to as "user agent" from this point forward. In a wider contest, the user can be a citizen, a company with devices with IoT, an organization, a city worker, a mobile app on behalf of a person, or more generally any entity that has to share a service with other peers in a smart city. In the real world, users are supposed to be told to use always the lowest priority, unless there were really unusual and urgent situations. In this case, they could use several high priority levels, but they should only use the highest one in life-death situations or the ones that could imply serious health risks.

The framework allows researchers to define different kinds of user agents by implementing the operations about (a) deciding when to ask a service and (b) selecting the priority. We defined two kinds of users for illustrating the current approach. The selfish user agents selected the highest priority. The default honest user agents mostly select the lowest priority except in a few cases that represent actual urgent matters. Both user agent kinds applied the mechanism for taking nondeterministic decisions proposed by TABSAOND (a technique for developing ABS apps and online tools with nondeterministic decisions) [20].

In each service request, an agent transforms the priority self-reported by the user into a priority that can be fairly compared with the priorities of other users. This agent would be actually software in the real world, and this approach simulates this with software agents called "priority normalizer agents" (also referred to as "priority agents"). Each priority

agent will be assigned to each citizen or any other kind of entity that is represented by the user agent. The priority normalizer agent is the one that would be extended to test different self-reported policies, and this can register all the previous priorities requested by its corresponding user.

The "resource manager agent" (also denoted as "manager agent") is in charge of assigning the services faster to the ones with the highest priorities. In order to avoid starvation of some users, the manager agent has to continuously attend all the requests, even if the higher priority ones are attended faster. For this purpose, this agent has several queues. Here, we illustrate the approach with three queues due to the selected number of priority levels, but it could be generalized for a different number of queues.

The next subsections further describe the most relevant aspects of respectively each kind of agents. In addition, Section 2.1.4 introduces the user interface (UI) of the simulator.

2.1.1. Behaviors of User Agents. In order to facilitate the definition of different user agents, firstly we defined a basic user agent called "stochastic user agent" that takes nondeterministic decisions about when to ask the services and which priority to report. This agent takes these decisions based on probabilities for the different cases, by applying TABSAOND. This agent was implemented to serve as a common basis for defining usual user behaviors in a simple way by just changing certain probabilities.

Regarding the decisions about when asking services, we defined a threshold that determines a timeline spot in which the agent is considered to go from novel user to experienced user. It defines different probabilities for deciding whether to ask the service or not, regarding whether it is a novel or an experienced user. Regarding the probability, in each simulated iteration it takes the decision with the following formula:

$$d_a = \begin{cases} \text{yes}, & \text{if } r \leq p_{a,e} \\ \text{no}, & \text{otherwise}, \end{cases} \tag{1}$$

where d_a is the decision about whether to ask a service, r is a random number in the interval $[0, 1]$, generated by a generic random number generator library of Unity engine, and $p_{a,e}$ is the probability of asking a service regarding the e experience of use (which can be either beginner or experienced user).

The calculation of $p_{a,e}$ is done by firstly distinguishing whether the user is a beginner or not by comparing their time using the service with a certain threshold. Then, $p_{a,e}$ is set to a specific probability value if the user is a beginner or is set to a different value otherwise. The selection of these two probability values can be established in a calibration phase of the system, for example, following the technique for Automatically Training ABS (ATABS) [21].

The beginner user is normally assumed to ask for more services, while experienced users may need a less frequent use of the service. In addition, with this model, the system can use a high amount of requests filling the queues of the different priority levels. Then, the users ask for a normal amount of requests, but with the queues already filled. In this way, we can properly compare the different priority policies.

After this, if the user agent requests a service, it selects the priority based on a distribution of probabilities. As a particular case of the suggestions of TABSAOND, it applies the following formula for the three priority levels, although this could be extended for a higher number of levels:

$$d_p = \begin{cases} 0, & \text{if } r \leq p_{p,0} \\ 1, & \text{if } p_{p,0} < r \leq \left(p_{p,0} + p_{p,1}\right) \\ 2, & \text{otherwise}, \end{cases} \tag{2}$$

where d_p is the decision of which priority level is assigned, r has the same meaning as in the previous equation, and $p_{p,0}$ and $p_{p,1}$ are the probabilities of respectively selecting the zero and one levels of priority. The probability of selecting the two levels (i.e., the lowest one) is the subtraction of these probabilities from one. The $p_{p,0}$ and $p_{p,1}$ probability values are normally calibrated when conforming the desired behavior of a specific kind of user agents.

The difference between the honest and the selfish agent is that the former asks the high priority levels with frequencies when they actually needed and the selfish agent mainly asks the highest level. In particular, the honest user is represented with the default stochastic user agent in which the probabilities are $p_{p,0} = 10\%$ and $p_{p,1} = 10\%$. The selfish user agent is defined as an extension of the default stochastic user agent in which the probabilities are $p_{p,0} = 98\%$ and $p_{p,1} = 1\%$.

Different kinds of selfish agents could be implemented for testing new security attacks and misuses in the requests of services. For example, new kinds of selfish agents can be implemented by extending the user agent class and implementing the "Live" method. For example, the denial of service (DoS) or the distributed DoS attacks could be implemented by calling many times to the "Ask Service" method from the Live method with the corresponding loop.

2.1.2. Normalization of Priority Agents. In the current approach, the normalizer priority agents represent a logical mandatory layer of the communication for requesting services in smart cities with self-reported priorities. These priority agents are in charge of adapting self-reported priorities into normalized priorities that can be meaningfully compared between different users.

In the current framework, one can easily define new strategies for normalizing self-reported priorities, by extending the priority agent (i.e., by extending the corresponding object-oriented class), and overriding just the "Normalize" method.

In order to associate the priorities and the users that reported these, each priority agent is in charge of managing only the petitions of one user agent. In this way, the priority agent can record its activities in their fields (i.e., object-oriented attributes) to consider the history of each user agent. Since the framework is implemented with PEABS, the priority agents could even communicate each other by accessing all the agents of simulation of the corresponding type.

To illustrate the current approach, we have defined two strategies, denoted, respectively, as "smart" and "simple" from this point forward. The smart strategy is based on trusting the self-reported priorities of users at the beginning, but it also records the history of the priorities that each user has assigned to its requests. Each priority agent calculates the average of the priorities requested by the user. If the average surpasses or is equal to certain value (i.e., most of the requests had the lowest priority), then it trusts their self-reported priority and forwards the petition with this priority. If the average priority is below that threshold (i.e., the agent assigns high priorities frequently), then the priority agent interprets that the user is misusing the priority system, and it forwards its petition with the lowest priority level. This can be formalized with the following formula:

$$x_N = \begin{cases} x_{\max}, & \overline{X_h} >= x_t \\ x_{\text{in}}, & \text{otherwise,} \end{cases} \quad (3)$$

where x_N is the outputted normalized priority, x_t is the threshold for average priority in the history, X_h is the history of priorities, $x_{\max}$ is the maximum number that represents the lowest priority level, and x_{in} is the priority reported by the user agent.

Since the framework is prepared to support different numbers of self-reported priority levels, the threshold is calculated from another inverted threshold with the following formula:

$$x_t = x_{\max} - x_{t,i}, \quad (4)$$

where $x_{t,i}$ is the original inverted threshold and x_t and $x_{\max}$ have the same meaning as before.

It is worth mentioning that the proposed ABS could be easily adapted to support other ranges of priorities. In particular, the internal parameter "max priority" is represented as a property of the "Params" class in the source code and can be set to any value. For example, it could be set to 5 for representing a 5-star system. The range of normalized priorities can be established with the "max norm. priority" property of the same class. The inversion of the priorities for setting high priorities with high numbers (e.g., for being 5 stars the maximum priority) can be easily implemented by changing the order of the corresponding loop in the Live method of the manager agent.

2.1.3. Resource Manager Agent. The resource manager agent is in charge of managing the requests of users, for attending them considering both their chronological order and their priorities. It uses the priorities normalized by the corresponding strategy mechanism through the normalizer priority agents.

The resource manager agent has a queue for each priority level. Each request is added to the queue of its priority level. The request is also added to the queues of the lower priority levels, to guarantee that a request is not attended later than any subsequent request of a lower level. Later, when a request is attended, this is removed from all the queues in which it was added.

In order to avoid the starvation of the requests of the lower priorities, all the queues are attended in all the simulation iterations. However, to guarantee a lower waiting time in urgent requests, the number of attended requests is different for each priority level, and these amounts are configurable. For example, in the current experiments, in each loop, the manager attended three requests from the zero-level queue, two requests from the one-level queue, and just one from the two-level queue, if all the queues had enough requests.

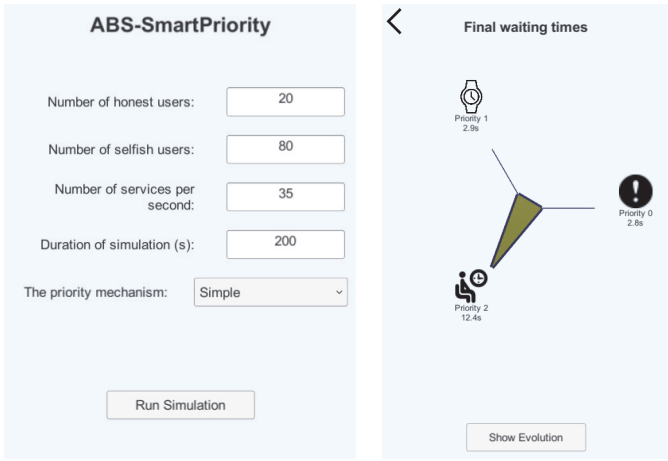
The resource manager agent knows the number of services that it can assign in each simulation iteration, due to the existing limited resources for attending the corresponding service in a smart city. Thus, in each iteration, it continues looping attending requests in the mentioned order, until the number of available assignments is reached or all queues are empty. Notice that if the queue of a priority level is empty, then the possible number of assignments is distributed among the other different priority levels.

2.1.4. User Interface. The UI of ABS-SmartPriority allows users to configure simulations for testing the different techniques for managing self-reported priorities in smart cities. Figure 1(a) shows the screen of the UI for entering the input parameters of the simulations. First, the user can introduce the numbers of respectively the honest users and the selfish users. The input screen also allows one to set the speed in which requests can be attended. This is set as the number of service requests that can be attended per second (each second is simulated with an iteration). The users of this app can also set the duration of the time that is simulated. Finally, the users can select a priority mechanism from a dropdown list in the UI. The user can press the "Run Simulation" button to start the simulation.

Once the simulation is executed, the application shows the final results with a starplot like the one of the execution example of Figure 1(b). Each end of the starplot is associated with a priority level that is shown with a label. An image is also displayed for each priority level to facilitate the understanding of the final results at a glance. Each end represents the average time that a fair observer user agent has waited when requesting services with each priority level.

The user can also press on the "Show Evolution" button to observe the evolution of average waiting times. The chart of this screen uses a different color for marking the waiting time of each priority level. The abscissa axis represents the simulated time, while the ordinates axis represents the waiting time. Section 3 will present examples of simulation evolutions.

2.2. Experimental Method. One of the main goals of the experimentation was to simulate different scenarios and strategies with the presented ABS and its framework, to determine whether this approach is useful for testing different strategies for fairly managing self-reported priorities. Another goal was to compare different ways of managing self-reported priorities in smart cities, determining whether the current approach can find statistically significant differences in their outcomes.



(a) Screen for the input parameters

(b) Screen with final simulation results

Figure 1: UI of ABS-SmartPriority.

Firstly, we defined two different kinds of citizens with ABS-SmartPriority. The first kind of citizens used the priority service properly assigning the lowest level of priority to the common operations and only used the high levels of priorities for urgent matters. The second kind of citizens was programmed to behave as selfish and tried to unfairly take advantage of the priority system by assigning a high priority in most requests including the ones related to common and daily situations.

Two strategies were defined for managing the self-reported priorities. One always trusted the self-reported priorities and forwarded them to the system. This strategy was used as the control mechanism of this experiment. The second strategy kept track of the users, controlling and normalizing the self-reported priorities by assigning the lowest priority level to the users that had abused the system. This abuse was detected from the excessive use of the high priority levels in their past history.

An observer agent was included to monitor the simulation. This agent impersonated an honest citizen who commonly asked low-priority requests except for infrequent urgent matters. This agent recorded the times that it waited for getting the different services for each priority level.

The ABS was executed respectively with the different strategies for managing the self-reported priorities. The results were compared to determine which policy provides a faster response for attending honest users.

The system was executed 100 times for each strategy to avoid bias due to the nondeterministic behaviors. The results were compared with a statistical test and the difference was measured with the corresponding effect size.

3. Results and Discussions

In the experimentation, we compared the different mechanisms for managing self-reported priorities, which are referred to respectively as the simple and smart strategies.

Table 1: Input parameters used in the experimentation.

Parameter	Value
Number of honest users	20
Number of selfish users	80
Number of services attended per second	35
Duration of the simulation (s)	200

ABS-SmartPriority was executed with same input parameters for each priority mechanism, and Table 1 shows these common input parameters.

In the system, we introduced an observer user agent that asked services with different priorities using a common honest priority probability distribution. This agent recorded all the times in which their service requests were attended. The simulator uses the average response times of these requests as the indicators of the QoS considering different priorities.

As mentioned in the experimental method, the simulator was executed 100 times for each priority mechanism. Table 2 presents the means of these simulations with the standard deviations (SD) between parentheses. This table also includes the differences of means and the percentage reductions. The latter ones are the percentages that the differences represent with respect to the value of the control mechanism (i.e., the simple strategy). One can observe that the smart technique reduces considerably the waiting time in the highest priority levels (i.e., levels zero and one) with reductions over 93%. In the lowest priority level, it achieves also reductions of waiting times with a reduction about 55%. Thus, this technique improves mostly the reduction of urgent requests, but it also reduces the waiting time of the daily requests, from the perspective of honest users.

This analysis has been performed from the perception and measurement of the experience of a honest user (i.e., the observer agent). It is worth mentioning that the reduction

Table 2: Comparison of waiting time means for the different priority levels between the different priority techniques.

	Priority 0	Priority 1	Priority 2
Simple waiting time (s)	3.218 (0.742)	3.004 (0.653)	10.610 (1.772)
Smart waiting time (s)	0.213 (0.500)	0.161 (0.376)	4.739 (0.634)
Diff. means (s)	3.005	2.843	5.871
Percentage reduction (%)	93.37	94.65	55.34

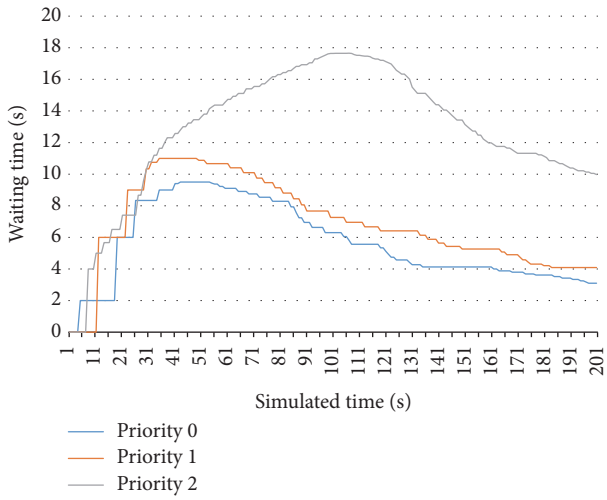


Figure 2: Results with the simple priority mechanism.

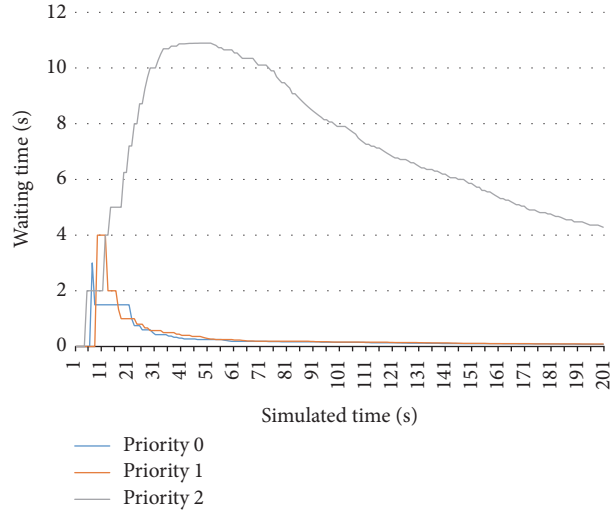


Figure 3: Results with the smart priority mechanism.

has been achieved by normalizing the self-reported priorities of selfish user agent, reassigning their priorities. Thus, the selfish users must wait as much as the other agents, letting the possibility of attending fast to actually urgent requests. From the selfish users' viewpoint, with the smart priority mechanism, they cannot take advantage of the priority system to be attended always fast with the highest priority. This is the reason why all the waiting times were reduced for honest users, although all the experiments used the same speed for providing smart city services.

Figure 2 shows the simulation outcomes of one of the executions with the simple priority mechanism, by showing its evolution. One can observe that the lowest priority level (i.e., priority 2) obtained higher response times (in average 12.8 s) than the other two priority levels (priorities 1 and 0 with, respectively, averages of 6.9 s and 5.7 s). This difference can also be noticed in the maximum waiting times. In the case of the lowest priority level the maximum was 17.7 s, while the higher levels spent at most 11.0 s and 9.5 s waiting for, respectively, priorities 1 and 0. In this way, one can observe that the resource manager agent worked properly, as even with the simple priority policy and with most of the petitions with the highest priority, the queues worked and maintained the highest priority levels with lower response times than the lowest priority.

Though, the highest priority level was still not so fast with an average of 5.7 s and a SD of 2.5 s, probably due to the misuse of the highest priority level on behalf of the selfish user agents, which was not controlled by the simple strategy.

In the evolution chart, one can also observe the intended effect of the user agents starting with high amounts of requests and then after a while lowering these amounts. This was reflected in the steep increase of waiting times, and then the later slow decrease of these.

Figure 3 shows an example of the results of a simulation of the smart strategy for managing self-reported priorities, by presenting the simulation evolution. The resource agent still maintained the highest priority levels with lower waiting times (in both cases with an average of 0.3 s) in comparison to the lowest priority (with an average of 7.1 s). This difference is also reflected in the maximum waiting times, which was 10.9 s for the lowest priority and 4.0 s and 3.0 s for, respectively, priorities 1 and 0. However, the difference was that the smart strategy detected the misuse of the highest priority levels on behalf of the selfish user agents and did not let them abuse these levels. In this manner, in the few times that the honest observer agent needed an urgent service, it got the service much faster. This is confirmed by the much lower response times of priority levels 1 and 0, which were 0.3 s in both cases in comparison to the simple strategy (previously shown in Figure 2) that obtained, respectively, 6.9 s and 5.7 s. At the end of the simulation, the queues were emptied and its use could be considered as the normal one. At the end of the simulation, one can observe that the two highest priorities were almost immediately attended, with waiting times of 0.08 s and 0.09 s, in comparison to the lowest priority level with a waiting time of 4.3 s.

Table 3: Robust tests of equality of means.

	Statistic[a]	df1	df2	Sig.
Priority 0				
Welch	1127.361	1	173.420	.000
Brown-Forsythe	1127.361	1	173.420	.000
Priority 1				
Welch	1423.749	1	158.261	.000
Brown-Forsythe	1423.749	1	158.261	.000
Priority 2				
Welch	973.528	1	123.955	.000
Brown-Forsythe	973.528	1	123.955	.000

[a] Asymptotically F distributed.

Table 4: Cohen's d effect size.

	Priority 0	Priority 1	Priority 2
Cohen's d	4.75	5.34	4.41

We applied Welch's t-test to determine whether the differences of means were statistically significant considering the 100 simulations of each strategy. We selected this test as it is robust for unequal variances. We also applied the Brown-Forsythe test for testing the equality of means, as this test is also robust for unequal variances. Table 3 shows the results of both tests. As one can observe in the results, both tests determined that the differences of means were very significant (considering a .001 significance value) for all the priority levels.

In order to meaningfully measure the differences between the waiting time means, we calculated Cohen's d effect sizes between the simple and smart strategies for each priority level. Table 4 shows the resulting Cohen's d effect sizes. According to Cohen's guidelines [22] and the later interpretations proposed by Rosenthal [23], all the effect sizes can be considered very large, as they all surpassed the threshold of 1.3.

Figure 4 visually shows the differences of waiting times in the highest priority level (i.e., the level zero) with a boxplot. One can observe that the smart priority mechanism substantially decreased the means, considering that these differences were much higher than the variances. Though, the smart priority mechanism also presented some outliers, probably due to the actual coincidences of urgent requests from different honest user agents. On the contrary, the simple mechanism strategy did not present outliers but the mean was quite higher. The reason probably was that the services were overloaded with all the high priority requests of selfish user agents and the users waited a regular high amount of time.

Figure 5 graphically presents the differences of waiting times in the one-level priority. In this case, the smart priority strategy also greatly reduced the means considering the variances. The existence of outliers only in the smart strategy may be due to a similar reason as in the previous case.

Finally, Figure 6 shows the boxplot for the lowest priority level. It is worth mentioning that in the viewpoint of honest users the waiting time reduced also in this level in comparing

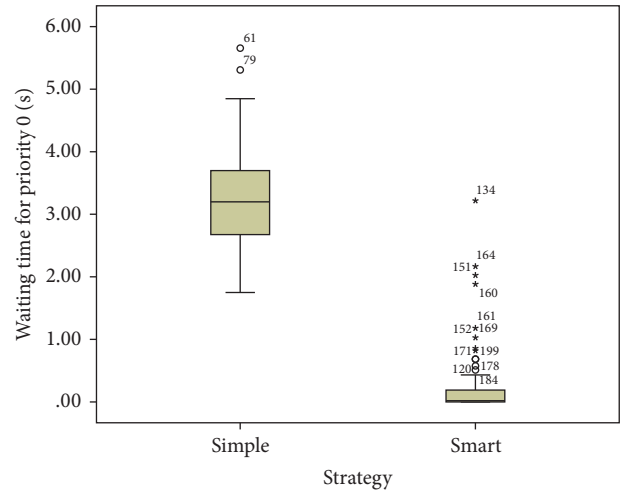


Figure 4: Comparison of waiting times of priority 0 between the simple and smart strategies.

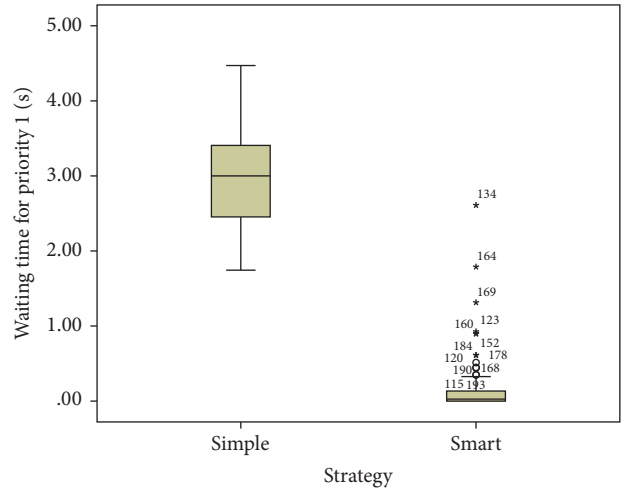


Figure 5: Comparison of waiting times of priority 1 between the simple and smart strategies.

when using the simple strategy. The reason is that the smart strategy stops the abuse of the high priority level. In the simple strategy, three selfish agents were normally attended every time an honest user was attended, since the former ones commonly used the highest priority level. This hindered the fair waiting time in the lowest level. In this boxplot, one can observe that the smart priority did not present outliers, probably since the requests of services of low-priority were more regular. Remember that honest agents mostly asked low-priority services and that the requests of selfish user agents were normalized as low-priority ones by the smart priority agents.

On the whole, ABS-SmartPriority has allowed us to properly simulate different self-reported priority policies that presented statistically significant differences in the outcomes.

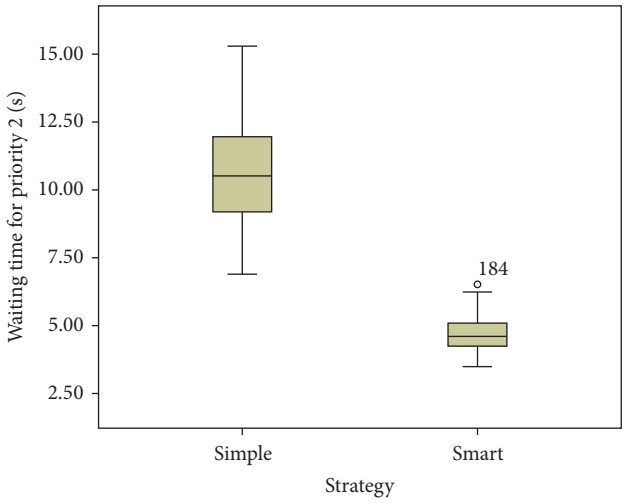


Figure 6: Comparison of waiting times of priority 2 when using a smart technique for controlling self-reported priorities.

The main contribution of the current work over works about priority mechanisms like the ones about waste collection [7] and hospital management [8] is that the proposed approach explores the management of self-report priorities by people with heterogeneous criteria and agendas in a nearly fair way. In addition, the current work improves the testbeds about smart priorities such as the one of Sanchez et al. [10] by incorporating the concept of self-reported priorities for increasing the performance of the use of smart city resources and services. Finally, the current work improves the previous works about respectively measuring the communication of MASs [24] and the integration of web services in agent-oriented methodologies [25] by presenting a MAS that manages the service requests considering priorities normalized from freely self-reported priorities of users.

4. Conclusions

The current work has presented an ABS for analyzing and comparing the repercussions of different strategies for managing self-reported priorities in smart cities. This simulator was illustrated by comparing two different strategies. The first one was based on tracking the history of each citizen to prevent them from unfairly overusing the high priority levels. The second one was a control mechanism that just used the priorities self-reported by the users. The experiments showed that the presented ABS-SmartPriority simulator and its underlying framework were useful to compare these strategies and detect statistically significant differences in the outcomes between these strategies. Therefore, the presented ABS can assist smart city engineers in preparing adequate priority protocols for managing the priorities self-reported by citizens through their apps and devices with IoT in order to fairly manage the smart city services.

In the future, this work is planned to be extended by experiencing a larger number of strategies for managing self-reported priorities. It will also be extended for defining strategies that manage several kinds of smart city services in different ways. We also plan to enhance this work by incorporating mechanisms that use trust and reputation techniques to interchange the reliabilities of the users that self-report their priorities. Furthermore, in the future the current approach will incorporate a feedback system in which the veracity of the reported emergencies may be checked after being attended.

The UI of the ABS is planned to incorporate more parameters about the experience of smart city users and their behaviors. In addition, the simulator will consider more features of users for determining their frequencies in requesting certain services. For example, more user agent types will be implemented with parameters such as the age of users, which is normally related to the need of healthcare services. The UI will be extended to configure these new user agent types with their parameters for the simulations.

Conflicts of Interest

The authors state that this work has not any conflicts of interest.

Acknowledgments

This work has been supported by the program "Estancias de movilidad en el extranjero José Castillejo para jóvenes doctores" funded by the Spanish Ministry of Education, Culture and Sport with Reference CAS17/00005. The authors also acknowledge support from "Universidad de Zaragoza," "Fundación Bancaria Ibercaja," and "Fundación CAI" in the "Programa Ibercaja-CAI de Estancias de Investigación" with Reference IT24/16. They acknowledge the research project "Construcción de un framework para agilizar el desarrollo de aplicaciones móviles en el ámbito de la salud" funded by University of Zaragoza and Foundation Ibercaja with Grant Reference JIUZ-2017-TEC-03. This work also acknowledges the research project "Desarrollo Colaborativo de Soluciones AAL" with Reference TIN2014-57028-R funded by the Spanish Ministry of Economy and Competitiveness. It has also been supported by "Organismo Autónomo Programas Educativos Europeos" with Reference 2013-1-CZ1-GRU06-14277. Finally, they acknowledge support from project "Sensores vestibles y tecnología móvil como apoyo en la formación y práctica de mindfulness: prototipo previo aplicado a bienestar" funded by University of Zaragoza with Grant no. UZ2017-TEC-02.

Supplementary Materials

This supplementary material contains the source code of the agent-based simulator of strategies for managing self-reported priorities in smart cities called ABS-SmartPriority. In this material, the file "SmartPriority.zip" of the project with the source code can be unzipped and opened with the Unity Editor 5.5.1. The material also contains a brief web-based documentation for presenting the simulator and indicating how to define new strategies. *(Supplementary Materials)*

References

[1] H. Yeh, "The effects of successful ICT-based smart city services: from citizens' perspectives," *Government Information Quarterly*, vol. 34, no. 3, pp. 556–565, 2017.

[2] H. B. Sta, "Quality and the efficiency of data in "Smart-Cities"," *Future Generation Computer Systems*, vol. 74, pp. 409–416, 2017.

[3] F. Vicentini, A. Giusti, A. Rovetta et al., "Sensorized waste collection container for content estimation and collection optimization," *Waste Management*, vol. 29, no. 5, pp. 1467–1472, 2009.

[4] L. Parra, S. Sendra, J. Lloret, and I. Bosch, "Development of a conductivity sensor for monitoring groundwater resources to optimize water management in smart city environments," *Sensors*, vol. 15, no. 9, pp. 20990–21015, 2015.

[5] T.-h. Kim, C. Ramos, and S. Mohammed, "Smart city and IoT," *Future Generation Computer Systems*, vol. 76, pp. 159–162, 2017.

[6] M. Yigit, V. C. Gungor, E. Fadel, L. Nassef, N. Akkari, and I. F. Akyildiz, "Channel-aware routing and priority-aware multichannel scheduling for WSN-based smart grid applications," *Journal of Network and Computer Applications*, vol. 71, pp. 50–58, 2016.

[7] T. Anagnostopoulos, K. Kolomvatsos, C. Anagnostopoulos, A. Zaslavsky, and S. Hadjiefthymiades, "Assessing dynamic models for high priority waste collection in smart cities," *The Journal of Systems and Software*, vol. 110, pp. 178–192, 2015.

[8] Y. de Harlez and R. Malagueño, "Examining the joint effects of strategic priorities, use of management control systems, and personal background on hospital performance," *Management Accounting Research*, vol. 30, pp. 2–17, 2016.

[9] G. Martínez-Cánovas, E. Del Val, V. Botti, P. Hernández, and M. Rebollo, "A formal model based on Game Theory for the analysis of cooperation in distributed service discovery," *Information Sciences*, vol. 326, pp. 59–70, 2016.

[10] L. Sanchez, L. Muñoz, J. A. Galache et al., "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, pp. 217–238, 2014.

[11] T. Anagnostopoulos, A. Zaslavsy, A. Medvedev, and S. Khoruzhnicov, "Top—k Query based dynamic scheduling for IoT-enabled smart city waste collection," in *Proceedings of the 16th IEEE International Conference on Mobile Data Management (MDM '15)*, pp. 50–55, Pittsburgh, Pa, USA, June 2015.

[12] I. García-Magariño, C. Gutiérrez, and R. Fuentes-Fernández, "The INGENIAS development kit: a practical application for crisis-management," *Bio-Inspired Systems: Computational and Ambient Intelligence*, vol. 5517, pp. 537–544, 2009.

[13] I. García-Magariño, J. J. Gómez-Sanz, and J. R. Pérez-Agüera, "A multi-agent based implementation of a Delphi process," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '08)*, vol. 3, pp. 1543–1546, International Foundation for Autonomous Agents and Multiagent Systems, Estoril, Portugal, May 2008.

[14] S. Hassan, M. Salgado, and J. Pavón, "Friendship dynamics: modelling social relationships through a fuzzy agent-based simulation," *Discrete Dynamics in Nature and Society*, vol. 2011, Article ID 765640, pp. 1–19, 2011.

[15] I. García-Magariño, "ABSTUR: An Agent-based Simulator for Tourist Urban Routes," *Expert Systems with Applications*, vol. 42, no. 12, pp. 5287–5302, 2015.

[16] J. P. Garcia-Sabater, J. Lloret, J. A. Marin-Garcia, and X. Puig-Bernabeu, "Coordinating a cooperative automotive manufacturing network - An agent-based model," in *Cooperative Design, Visualization, and Engineering*, vol. 6240 of *Lecture Notes in Computer Science*, pp. 231–238, 2010.

[17] I. García-Magariño, A. Gómez-Rodríguez, J. C. González-Moreno, and G. Palacios-Navarro, "PEABS: a process for developing efficient agent-based simulators," *Engineering Applications of Artificial Intelligence*, vol. 46, pp. 104–112, 2015.

[18] I. Garca-Magari, R. Lacuesta, and I. Garca-Magariño, "Source code of ABS-SmartPriority," 2017, Published as supplementary material of the current article.

[19] I. García-Magariño, M. Cossentino, and V. Seidita, "A metrics suite for evaluating agent-oriented architectures," in *Proceedings of the 25th Annual ACM Symposium on Applied Computing (SAC '10)*, pp. 912–919, Sierre, Switzerland, March 2010.

[20] I. García-Magariño, G. Palacios-Navarro, and R. Lacuesta, "TABSAOND: A technique for developing agent-based simulation apps and online tools with nondeterministic decisions," *Simulation Modelling Practice and Theory*, vol. 77, pp. 84–107, 2017.

[21] I. García-Magariño and G. Palacios-Navarro, "ATABS: a technique for automatically training agent-based simulators," *Simulation Modelling Practice and Theory*, vol. 66, pp. 174–192, 2016.

[22] J. Cohen, *Statistical Power Analysis for The Behavioral Sciences*, Lawrence Earlbaum Associates, Hillsdale, NJ, USA, 2nd edition, 1988.

[23] J. A. Rosenthal, "Qualitative descriptors of strength of association and effect size," *Journal of Social Service Research*, vol. 21, no. 4, pp. 37–59, 1996.

[24] C. Gutiérrez Cosio and I. García Magariño, "A metrics suite for the communication of multi-agent systems," *Journal of Physical Agents*, vol. 3, no. 2, pp. 7–14, 2009.

[25] R. Fuentes-Fernández, I. Garca-Magariño, J. J. Gómez-Sanz, and J. Pavón, "Integration of web services in an agent-oriented methodology," *International Transactions on Systems Science and Applications*, vol. 3, pp. 145–161, 2007.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 8789513, 15 pages
https://doi.org/10.1155/2017/8789513

Research Article

Handover Based IMS Registration Scheme for Next Generation Mobile Networks

Shireen Tahira,[1] Muhammad Sher,[1] Ata Ullah,[2] Muhammad Imran,[3] and Athanasios V. Vasilakos[4]

[1]*Department of Computer Science and Software Engineering, International Islamic University, Islamabad 44000, Pakistan*
[2]*Department of Computer Science, National University of Modern Languages, Islamabad 44000, Pakistan*
[3]*College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia*
[4]*Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, 971 87 Luleå, Sweden*

Correspondence should be addressed to Shireen Tahira; shireentahira381@gmail.com

Received 27 May 2017; Revised 2 August 2017; Accepted 7 August 2017; Published 27 September 2017

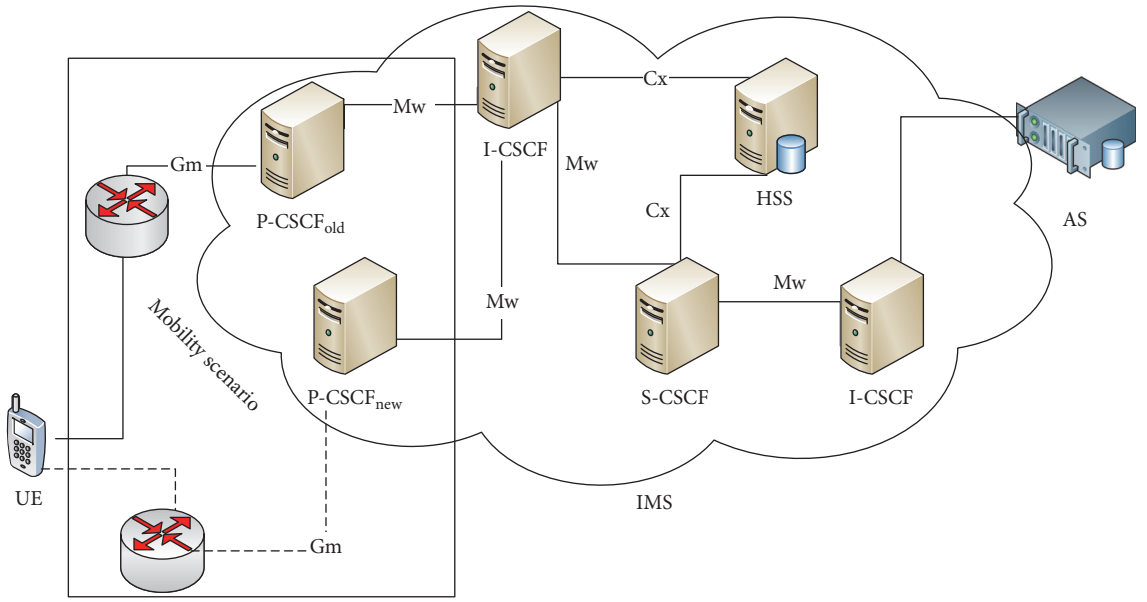
Academic Editor: Syed Hassan Ahmed

Next generation mobile networks aim to provide faster speed and more capacity along with energy efficiency to support video streaming and massive data sharing in social and communication networks. In these networks, user equipment has to register with IP Multimedia Subsystem (IMS) which promises quality of service to the mobile users that frequently move across different access networks. After each handover caused due to mobility, IMS provides IPSec Security Association establishment and authentication phases. The main issue is that unnecessary reregistration after every handover results in latency and communication overhead. To tackle these issues, this paper presents a lightweight Fast IMS Mobility (FIM) registration scheme that avoids unnecessary conventional registration phases such as security associations, authentication, and authorization. FIM maintains a flag to avoid deregistration and sends a subsequent message to provide necessary parameters to IMS servers after mobility. It also handles the change of IP address for user equipment and transferring the security associations from old to new servers. We have validated the performance of FIM by developing a testbed consisting of IMS servers and user equipment. The experimental results demonstrate the performance supremacy of FIM. It reduces media disruption time, number of messages, and packet loss up to 67%, 100%, and 61%, respectively, as compared to preliminaries.

1. Introduction

Next generation mobile networks (NGMNs) provide faster speed and wide capacity for outdoor and indoor multimedia based applications in affordable rates for supporting large set of customers. It can benefit this era where social networks are sharing huge data and video transfer across the world. NGMN consists of two layers [1] including transport layer and service control layer where the former manages the connectivity and transmission using WiMAX, UMTS, and LTE and the latter utilizes IMS servers for registration, integration, charging, and policy enforcement as per 3rd Generation Partnership Project (3GPP). User equipment (UE) connects through transport layer and then registers with IP Multimedia Subsystem (IMS) using service control layer in most access networks such as LTE/LTE-A. UE is mostly connected to the Internet and can establish audio/video calls anywhere and anytime. During traveling, UE can move out of a current network range and may join another neighboring network by performing handover to support uninterruptible call sessions. To provide seamless vertical handover across different heterogeneous NGMNs, various solutions using IMS were proposed in the literature [2–5].

During vertical handover, UE has to deregister with the old network and reregister with the new in order to continue uninterrupted service, which also requires obtaining a new IP address from the latter and reregistering with IMS. Moreover, IPSec Security Associations (SAs) have to be established again between UE and Proxy-Call Session Control Function (P-CSCF) to secure the mutual communications. During handover scenarios, UE changes its P-CSCF which is the first entry point for reregistration at IMS as illustrated in Figure 1.



Mobility occurs when UE connects to new AR and new P-CSCF

FIGURE 1: Reregistration at IMS in handover scenario.

UE exchanges messages over *Gm* interface to initiate registration process with P-CSCF that further contacts Interrogating-Call Session Control Function (I-CSCF) by using *Mw* interface. I-CSCF is responsible to get the address of Serving-Call Session Control Function (S-CSCF) that is to be assigned to the UE. I-CSCF retrieves the address using *Cx* interface to contact Home Subscriber Server (HSS) that maintains subscription related information of UE in a database. Application Server (AS) hosts and executes the services offered by IMS. UE has to go through this phase each time it connects to IMS that involves four messages from UE to IMS. It increases latency and disruption time that affects QoS [6], increases number of messages [7], and causes more packet loss [8].

Some of the existing reregistration schemes proposed in the literature [7, 9, 10] strived to reduce the number of messages and media disruption time (MDT), that is, time during which service is unavailable due to handover. Most of the schemes address delay in MDT by suggesting either integrating new protocols [2, 4, 5, 9–12] or context transfer from old servers to the new [5, 7]. Preprocessing is suggested by [11–13] in order to save time. The issue of changing the IP address of UE during handover is also addressed in [9] but not within Session Initiation Protocol (SIP) capabilities. SIP [14] is the main protocol of IMS for creating, modifying, and terminating sessions between UE and Corresponding Node (CN). Minimizing number of messages is addressed in [7, 9]. Mobile IPv6 (MIPv6) [15] and Fast MIPv6 (FMIPv6) in IMS solve mobility and latency problems. Media Independent Handover (MIH) [16] is also exploited to get mobility information of target network for preregistration in many schemes in order to reduce latency of layer 2 mobility. But none of the above methods provided the solution for authentication and authorization of entities needed in IMS registration after handover. In 3GPP specification, a new entity named service centralization and continuity application server is introduced for service continuity by maintaining long-term session between UE and CN on a new IP address. It ensures service continuity but it causes delay during registration phase.

The main problem during registration is that it undergoes latency in MDT, loss of packets, and more number of commands/messages. From literature, we have identified that there is a need to optimize the registration phase with less communication overhead. It should also cater transfer of IPSec SAs to provide authentication and authorization. The scheme should be generic that can be used with any NGMNs handovers between either LTE and WiMAX or LTE and 5G. Until now cross-layer schemes are proposed that reduce the registration delay but only by utilizing either layer 2 or layer 3 protocols. This paper presents a Fast IMS Mobility (FIM) scheme that maintains a flag to ensure that no deregistration in IMS occurs if flag is enabled during handover. During handover, UE does not intentionally terminate from AN and keeps a session with CN to get registered in IMS again. We have recommended that, after layer 3 handover and discovery of new P-CSCF, UE should transmit a new subsequent request for handover to S-CSCF for replacing old IP address of UE with new one. It also initiates transferring IPSec SAs to new P-CSCF. In this way, four messages of REGISTER request are reduced to two messages. Similarly it reduces number of commands by eliminating the need of network initiated deregistration in order to reduce communication overhead. We have set up a testbed for IMS along with supported servers and UE to implement and validate the handover process. Results prove the dominance of FIM as compared to IMS handover process and MIP based handover schemes in terms of minimizing the MDT, packet loss, and number of commands.

The rest of the paper is organized as follows. Section 2 describes the system model and related schemes for handover

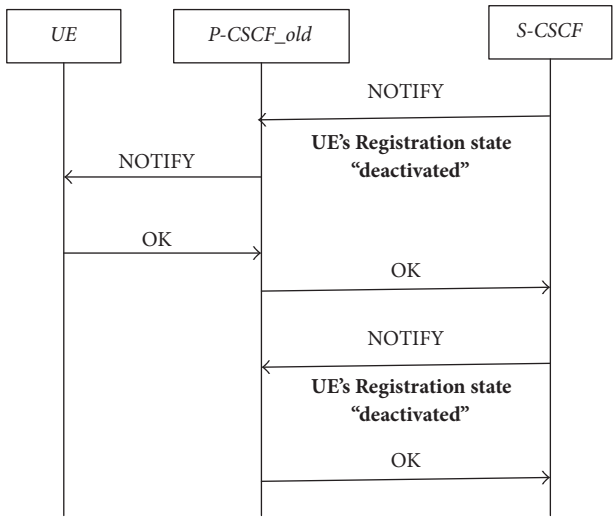


Figure 2: Network initiated deregistration in IMS.

in NGMN using IMS are discussed in Section 3. The proposed scheme is described in Section 4. Section 5 elucidates testbed implementation, results, and analysis. Finally, our work is concluded in Section 6.

2. System Model

We consider an IMS model where the main entities are CSCF servers, its database, and application server [17]. Figure 2 illustrates the flow of steps during network initiated deregistration in IMS during handover. After disconnection of UE from AN, the S-CSCF sends a NOTIFY message to P-CSCF_Old that forwards it to UE to inform it about deactivation of its registration. UE acknowledges it by sending OK response. S-CSCF also notifies old P-CSCF about deactivation of UE and old P-CSCF responds by OK message.

On connecting to a new AN, UE gets registered to IMS via new P-CSCF. Initially, a REGISTER request is sent to new P-CSCF who forwards it to I-CSCF that shares user authentication request (UAR) and answers UAA with HSS. After that I-CSCF transmits the REGISTER request to S-CSCF that exchanges message authentication request (MAR) and answers (MAA) with HSS and then replies with the 401 unauthorized message. Upon receiving 401, UE again transmits a REGISTER message to S-CSCF through P-CSCF and I-CSCF. S-CSCF exchanges server authentication request (SAR) and answers SAA with HSS and then sends OK message as shown in Figure 3. IPSec SAs are also established between UE and P-CSCF to ensure that these messages are exchanged between authorized UE and servers. During registration, the SIP headers including *Via*, *Route*, *To*, *Contact*, *Path*, and *Service-Route* are also used. Every time UE gets connected to new AN, the whole phase of registration is carried out that increases the delay in MDT and number of commands. Delay in MDT causes degraded QoS because of interruption in data flow and loss of packets. We assume that an intrusion detection system is functional to guard against security breaches and failure of messages by adversaries. It is also assumed that neighboring

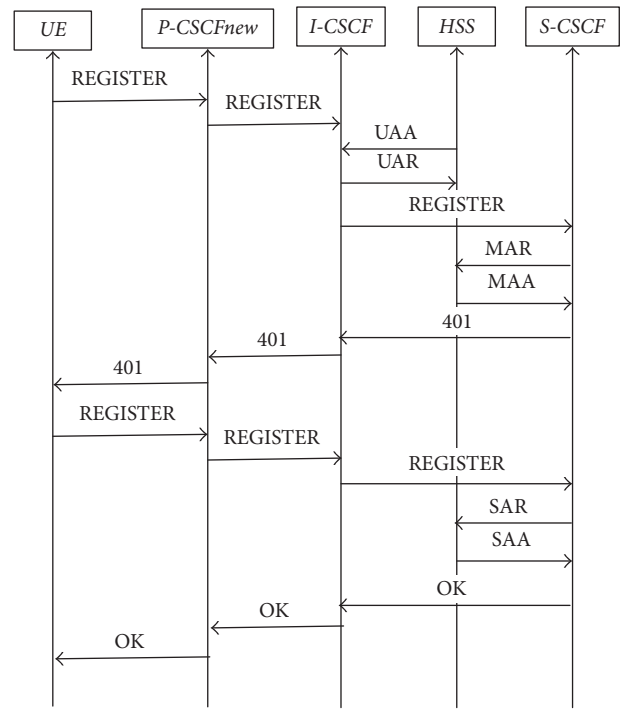


Figure 3: UE's registration at IMS.

AN is trustworthy to transfer session credential during handover and change of IP address is protected from IP spoofing.

3. Related Work

In this section, we discuss different schemes for handover scenarios and related functionalities by using IMS. Related schemes are discussed under different categories including handover schemes with less MDT where the target is to reduce the MDT during the handover process. It is further subdivided into new protocol inclusion based schemes, context transfer, preprocessing, simultaneous processing, and minimizing number of commands based schemes. In addition to these, change in IP address based schemes and packet loss evaluation based schemes are also discussed in the following section.

3.1. Handover Schemes with Less MDT. In literature, we have studied many techniques that are proposed to reduce MDT. We have discussed related schemes under different categories including integration of new protocols, context transfer, preprocessing, and simultaneous processing.

3.1.1. New Protocol Inclusion in IMS Based Schemes. These schemes recommend including new protocols in IMS for reducing MDT. FMIPv6 is developed for reducing layer 3 delay [18] that is recommended to be integrated with IMS to reduce handover delay [10, 11]. Figure 4 shows the MIP based mobility solution for IMS handover where two messages are exchanged between UE and old Access Router (AR-old) for router solicitation/advertisement. Two messages are exchanged between UE and Home Agent (HA) for Binding

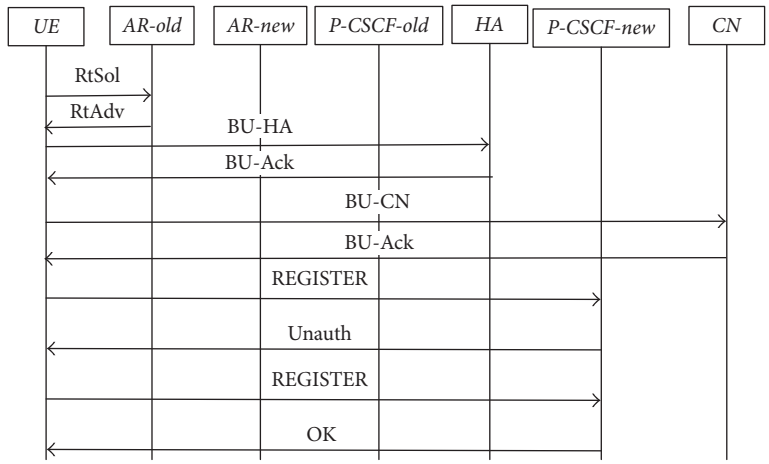


Figure 4: MIP-IMS handover.

Update (BU) and BU Acknowledgment (BU-Ack). Similarly two messages are exchanged between UE and CN for BU and BU-Ack. Then standard registration phase is carried out via new P-CSCF. Thanh et al. [19] have presented a solution to integrate mSCTP with IMS with a feature of multihome. When handover occurs, UE establishes a new mSCTP connection with new network while the old connection with previous network is still active. A new SWITCH SIP method is proposed in order to quickly switch between these two networks. Once the second mSCTP connection is established, switching between two networks is done with less latency. But in this way application layer latency caused by handover is not minimized and the integration issues of MIPv6 in IMS are inherited in FMIPv6 as well [6].

A cross-layer architecture has been proposed for handover from LTE to WiMAX by integrating MIP and SIP protocols [2]. It uses evolved packet core as the core network and IMS to provide multimedia services and manage sessions. The delay (D) for handover from LTE to WiMAX is calculated using (1), where D_{MIP} is calculated using (2) by calculating the delay time of agent solicitation/advertisement for the Base Station (BS) of WiMAX and registration request and reply with HA to identify the new P-CSCF and S-CSCF. The time T_{UEtoCN} is calculated for BU and BU-Ack messages that are exchanged between UE and CN.

$$D = D_{\text{MIP}} + D_{\text{SIP}} + D_{\text{HSS}} \tag{1}$$

$$D_{\text{MIP}} = 2T_{\text{UEtoBS}} + 2T_{\text{UEtoHA}} + 2T_{\text{UEtoCN}}. \tag{2}$$

Equation (3) explores the calculation of D_{SIP} for non-cross-layer architecture that involves the delay time taken for the registration request and response messages between UE and S-CSCF. It also involves delay time for 3 Re-INVITE messages between UE and CN, that is, REINVITE, OK, and ACKNOWLEDGMENT. Equation (4) shows $D_{\text{SIP\_CL}}$ for cross-layer scenario.

$$D_{\text{SIP}} = 2T_{\text{UEtoS}} + 3T_{\text{UEtoCN}} \tag{3}$$

$$D_{\text{SIP\_CL}} = 2T_{\text{HAtoS}} + 3T_{\text{UEtoCN}}. \tag{4}$$

In this case, D_{HSS} is the delay consumed by S-CSCF to update HSS about new location of UE. In this scheme registration of UE with S-CSCF is shown in 2 messages but it does not explain the authentication, establishment of SAs with new P-CSCF, and integrity protection issues.

3.1.2. Context Transfer Based Schemes. Techniques proposed in [7, 9] transfer the context from old P-CSCF to new P-CSCF for reducing number of messages for registration and session establishment that leads to reduction in MDT. It also contains the information of the user and state of session that was being used before the UE changes its network. Thus it reduces number of messages and handover latency. Farahbakhsh and Movahhedinia [10] proposed two schemes for supporting seamless handover and two schemes for QoS parameter negotiation. These schemes are proposed to reduce latency by transferring the context from old servers to new servers. These techniques are either predictive (user knows in advance to which network it has to move) or reactive. These techniques use FMIPv6 to remove latency of handover. Equation (5) shows its predictive scheme where four messages are exchanged between UE and old AR ($4T_{\text{UEtoARo}}$). One message is sent to old P-CSCF from UE (T_{UEtoPold}). Two messages are exchanged between old AR and new AR ($2T_{\text{ARotoARn}}$). One message is sent from UE to new AR (T_{UEtoARn}). Two messages are exchanged between UE and HA ($2T_{\text{HA}}$). Four messages are exchanged between UE and CN ($4T_{\text{UEtoCN}}$). Two messages are exchanged between UE and new P-CSCF ($2T_{\text{UEtoPnew}}$). Equation (6) shows MDT_{RM} for reactive mode where T_{PoldtoS} is the message between old P-CSCF and S-CSCF.

$$\begin{aligned}\text{MDT} = {} & 4T_{\text{UEtoARo}} + T_{\text{UEtoPold}} + 2T_{\text{ARotoARn}} \\ & + T_{\text{UEtoARn}} + 2T_{\text{HA}} + 4T_{\text{UEtoCN}} \\ & + 2T_{\text{UEtoPnew}}\end{aligned} \tag{5}$$

$$\begin{aligned}\text{MDT}_{\text{RM}} = {} & 2T_{\text{UEtoARo}} + 3T_{\text{UEtoPnew}} + T_{\text{UEtoARn}} \\ & + 2T_{\text{ARotoARn}} + 3T_{\text{PoldtoPnew}} + 2T_{\text{PoldtoS}} \\ & + 2T_{\text{PoldtoS}} + 2T_{\text{HA}} + 4T_{\text{UEtoCN}}.\end{aligned} \tag{6}$$

3.1.3. Preprocessing Based Schemes. Few researchers introduced the approaches to do some kind of preprocessing before doing any handover to a new network. UE connects to a new network and then it rejects it for the reason that it does not provide required QoS and goes for the selection of a new network repeatedly till it finds the best suitable network. This process also adds up in MDT due to handover to new networks in the hunt of best network. Ito et al. [11] have proposed that MDT can be minimized by starting the service continuity procedure as soon as UE gets an IP address proactively due to FMIPv6. Buffering should be performed to minimize packet loss. If the network is not suitable one, then UE does not register in IMS with this network; rather UE searches for next available network. Nazari et al. [12] proposed a solution to minimize handover delay by eliminating the delay of registration in target network and IMS. With the help of MIH protocol [20], handover server enables the UE to get mobility information and then it preregisters with new network and IMS without being in coverage area. For preregistration in core network, authentication process should be done. For authentication process EAP (Extensible Authentication Protocol) is carried out. In this method, a handover between WiMAX and LTE is illustrated. For a handover to LTE, the security key used in the target BS cannot be constructed without knowing the BS's identity. Similarly in case of handover to WiMAX, master session key is generated and sent to target network but security key cannot be generated until identity of the BS is known. So the preregistration in the BS ($D_{\text{B-PR}}$) is done when UE enters in the coverage area. Therefore the delay of handover preparation is equal to time of registration in the BS only. MDT is calculated as Delay = $D_{(\text{B-PR})}$, where $D_{\text{B-PR}}$ is the preregistration in the BS.

Yang and Chen have proposed a QoS reservation model [21] that reserves resources in advance at neighboring IMS networks and manages mobility by using SIP multicast [22]. During handover between LTE and WiMAX, a cross-layer architecture [4, 5] recommends sending prior Re-INVITE to new AN while it is connected to old AN. During the reregistration phase, two messages are exchanged that contain IPSec SA parameters and UE's new IP address. The authors claimed that the total MDT in this scheme is only the IMS delay because switching to underlying link by UE is parallel to session setup. Moreover, the prior duplicate address detection [23] delay at layer 3 is also removed by using MIH services. For vertical handover between 3G and WiFi, many schemes are proposed to reduce delay by preregistration and pre-Re-INVITE [24]. Moon proposed an IPSec SA establishment method [3] in such schemes. It adds new SIP headers to resolve the IP address mismatch problem for IPSec SAs by indicating the destination IP address of UE in its SIP messages. Thus no detail is mentioned about how layer 3 mechanism reduces the overall MDT.

3.1.4. Simultaneous Processing Based Scheme. Another approach to reduce handover delay is to keep processing in old network while working on new network. Bellavista et al. have proposed a framework [25] where a module predicts the network on which UE is going to handover its connection. It keeps the flow on old network and also starts signaling on new network. Another module is to handle QoS issue on new network. In all these existing schemes, UE still has to get registered in IMS with new IP address and the registration/authentication process also runs once again that adds delay to MDT.

3.1.5. Minimizing Number of Commands Based Schemes. Larsen et al. proposed a scheme in order to reduce number of SIP messages for registration and invite (for session) in order to reduce MDT [7]. It proposed reusing the information about user and session states in CSCF servers. This technique claims to reduce number of messages from 15 to 4. It transfers the context from old P-CSCF to new P-CSCF, which contains the information of the user and state of session that was going on before user changed the network. It does not explain the methods of secure transfer of IPSec SAs to new P-CSCF, authentication, and authorization of UE.

3.2. Change of IP Address Based Schemes. Different schemes recommend integrating Mobile IP [9] in IMS as SIP does not provide a solution to change IP address of UE after handover. MIPv6 [15] is a layer 3 protocol for providing mobility solutions and reduces layer 3 mobility delay. It hides mobility of UE from upper layers. On changing AN, UE gains a Care of Address (CoA) that is saved in another entity Home Agent (HA). Communication is done by addressing Home Address (HoA) of UE. Message destined to UE goes to HA that finds the CoA of UE and forwards the messages to UE. Upper layers are unaware of the mobility of UE. MIP has integration issues that are discussed in [6].

3.3. Packet Loss Evaluation Based Schemes. Bagubali et al. [8] analyzed and evaluated the IMS based integration architecture proposed in [2] for WiMax/LTE handover. Cross-layer architecture is better than simple layered architecture by evaluating packet losses and other parameters [8]. Equation (7) is used by [26] to measure packet loss where Ti_{ad} is agent advertisement signal, G is downlink packet transmission rate, D is handoff delay, and Nm is the number of handovers during a session.

$$\text{Packet Loss} = \left[\left(\frac{1}{(2 * \text{Ti}_{\text{ad}})}\right) + D\right] * G * \text{Nm}. \quad (7)$$

4. Fast IMS Mobility (FIM) Scheme

This section discusses our proposed scheme that handles the handover and registration process when users are moving in a region. During mobility scenarios, UE needs to be transferred from one AN to the other without interrupting its session. Authentication and authorization of UE are also needed after handover so these are also catered in our scheme. Our scheme also gives a mechanism to transfer IPSec SAs from old P-CSCF to new P-CSCF in a secure manner. FIM scheme benefits all NGMN including 4G and 5G because it gives the solutions on application layer. The notations used in FIM are listed in the Notations.

In our scheme, the communication begins when UE transmits a proposed subsequent HANDOVER request to old

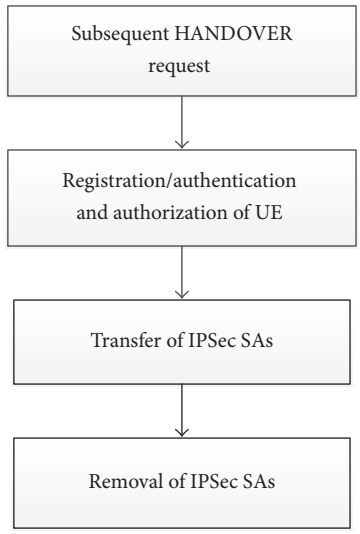


Figure 5: Phases of FIM scheme during handover.

HANDOVER sip:home1.fr.SIP/2.0
Via: SIP/2.0/UDP [5555::1:2:3:4];branch=0uetb
Route:sip:[5555::a:f:f:e];lr Max-Forwards: 70
From: sip:user@home1.fr;tag=pohja
To:sip:user@home1.f
Contact: sip:[5555::1:2:3:4};expire=600000
discovered-PCSCF:sip[6666::d:e:e:f]
TIPSecSAs:tls;q=0.2,IPSec3gpp;q=0.1;alg=hmac-sha-1-96;
spi-c=9865432;spi-s=8764321;port-c=8642;port-s=7531
Authorization:
Digest username="user1@home1.ims,
Response="083493483927jdhfjshfj"
Call-ID:ahedew23398fk
CSeq: 20 HANDOVER Content-Length:0

Box 1: Subsequent HANDOVER request message.

P-CSCF. HANDOVER request is sent instead of REGISTER request on the basis of status of "sessionContinued" flag in case it is enabled. It ensures that the UE and CN have not cancelled the session between them and the UE is disconnected from AN for a short period that could be the handover purpose. After that, UE's registration/authentication process begins at S-CSCF to proceed with the connectivity of new P-CSCF along with credentials. In the next phase, IPSec SAs are forwarded by the S-CSCF to new P-CSCF that exchanges the confirmation messages for the running call session. Finally, the old P-CSCF removes the IPSec SAs from itself as illustrated in four phases in Figure 5.

In the first phase, UE transmits a subsequent HANDOVER request with its new IP address in Contact and IP address of P-CSCF in discovered-PCSCF header as shown in Box 1. If the flag is not set, then REGISTER request will be sent to old-P-CSCF as illustrated in Pseudocode 1. UE's

IF sessionContinued == true then
Construct method = HANDOVER
Route "HANDOVER" to old P-CSCF
ELSE
Route "REGISTER" to old P-CSCF
ENDIF

Pseudocode 1: Session handling at UE.

disconnection from an AN leads to deregistration from IMS. Our scheme avoids deregistration due to *sessionContinued* flag. In this way, all the related data of UE's registration is not deleted from the IMS servers and UE. After layer 3 mobility, UE discovers a P-CSCF that is either new one or the same one at previous location. FIM sends HANDOVER request to S-CSCF via old P-CSCF instead of starting registration. The HANDOVER request utilizes a routing header, that is, *Service-Route*. In this way, P-CSCF does not need to contact I-CSCF for S-CSCF's address. So the I-CSCF does not contact HSS that results in no *Cx* commands for our scheme.

In the second phase, UE connects to a new AN and verifies that if the status of *sessionContinued* flag is enabled, then, instead of sending REGISTER request to IMS, UE constructs HANDOVER request. UE integrates its new IP address in Via and *Contact* headers. UE places IP address of new P-CSCF in a new proposed header *discovered-PCSCF* in HANDOVER message. UE forwards this HANDOVER request to old P-CSCF that forwards it to S-CSCF where new values from *Contact* and *discovered-PCSCF* headers are saved. IP address of new P-CSCF is copied to *path*. Now S-CSCF is able to forward request destined for UE towards new P-CSCF. In this way, UE gets registered in IMS. UE and HSS keep a shared secret and sequence number. In UE, it is saved in ISIM application on universal integrated circuit card (UICC). In the initial REGISTER request of standard IMS scheme, S-CSCF downloads authentication vectors (AVs) for UE. AVs contain random challenge (RAND) and expected response (XRES) along with integrity key (IK), cipher key (CK), and authentication token (AUTN). S-CSCF sends these AVs to UE in initial request's response, that is, 401. UE calculates RES with the help of shared secret and RAND. This RES is like a password [27] and used for authentication. In our proposed scheme, due to no deregistration, RES and XRES are not lost and used for authentication during HANDOVER phase. When S-CSCF receives HANDOVER, it compares the received RES with XRES and then saves the new IP address of UE and IP address of P-CSCF as illustrated in Pseudocode 2. In other scenarios, when the method is not HANDOVER, then complete registration process is performed including user RES verification in the reply of "Unauthorized" message. During registration, UE and P-CSCF establish IPSec SAs that ensures integrity protection/authorization. Our scheme sends HANDOVER request through old P-CSCF where IPSec SAs are saved in a header. By sending the request through old P-CSCF, the integrity of messages is ensured because UE and old P-CSCF have already established IPSec SAs between them. Thus it guarantees that messages are exchanged between

```
SCSCF receives request
IF method == "HANDOVER"
  IF RES == XRES & integrity-protected == true then
    IPaddress = Contact;
    P-CSCF_IPaddress = Path;
    IF P-CSCFold != P-CSCFnew then
     Route "HANDOVER" to P-CSCFnew
    ENDIF
  ELSEIF method == "REGISTER"
    IF RES != XRES then
     Create User-challenge( );
     route(Service-Routes);
     reply("401", "Unauthorized - Challenging UE");
    ELSEIF RES == XRES then
     Set -status == "200"
     Route "OK" to ICSCF
    ENDIF
  ENDIF
ENDIF
```

Pseudocode 2: Authentication at S-CSCF.

trusted entities. In this way, our scheme handles registration, authentication, and authorization in two messages.

In the third phase, the old P-CSCF receives message from UE and compares that if the received request is HANDOVER, then add IPSec SAs to HANDOVER message in TIPSecSAs header. It adds the IP address of new P-CSCF to *path* header and set *Integrity-Protected* to yes in HANDOVER request. Old P-CSCF does not send IPSec SAs to new P-CSCF directly because there is no IPSec SAs tunnel between these two entities. In our scheme, IPSec SAs are transferred from old P-CSCF to S-CSCF. S-CSCF first compares RES with XRES and then transfers the received IPSec SAs to new P-CSCF. Thus the transfer of IPSec SAs from old P-CSCF to new P-CSCF is done in a secure manner with authenticity. This is only done when the newly discovered P-CSCF is not the old one (before handover) because the transfer is only needed in that case. In this way, no new protocol is used to transfer IPSec SAs. No new commands are used for the transfer of IPSec SAs as it is done within the handover phase proposed by our scheme. In the fourth phase, S-CSCF transmits "OK" message to old P-CSCF. As in our scheme, there is no deregistration, so in order to delete IPSec SAs from old P-CSCF, the entity removes IPSec SAs from itself after sending "OK" message to UE as shown in Pseudocode 3. It is only needed when newly discovered P-CSCF is the new one. The registration after handover is presented in a visual manner in Figure 6 by sequentially exploring the above four phases.

In this way, we have reduced MDT between two entities after handover from one AN to another. MDT is confined by reducing delay and cost of the registration phase in IMS. Similarly we proposed reducing number of *Gm*, *Mw*, and *Cx* commands. Our proposed scheme handles authorization of UE and transfers IPSec SAs during registration phase. It handles all this with low latency in MDT as given in (8), where $E = \{\text{UE, PCSCF, ICSCF, HSS, SCSCF}\}$ is the set of entities, x is the number of messages between two entities, and $T(E_i, E_j)$ is delay time taken by the message between entities E_i and E_j, where $i = \{1, 2, 3, 4\}$ and $j = i + 1$.

$$\text{minimize MDT} = \sum x \cdot T\left(E_i, E_j\right) \tag{8}$$

For minimizing number of commands, if we consider $S = \{\text{PCSCF, ICSCF, SCSCF}\}$ is the set of IMS servers and $H = \{\text{HSS}\}$ is the set of database. Here y is the number of Mw commands, $i = \{1, 2\}$. Then number of Mw commands is needed to be minimized as given in equation (9). It calculates sum of Mw commands needed from P-CSCF to I-CSCF and from P-CSCF to S-CSCF. Equation (10) calculates sum of Cx commands needed from I-CSCF to HSS and from I-CSCF to HSS where z is the number of Cx commands.

$$Mw \text{ commands} = \sum y \cdot \left(S_i, S_{i+1}\right) \tag{9}$$

$$Cx \text{ commands} = \sum z \cdot \left(S_{i+1}, H\right). \tag{10}$$

$U = \{\text{UE}\}$ and k is the number of *Gm* commands; then the minimization of *Gm* commands is given in (11) that calculates the sum of Gm commands needed from P-CSCF to UE.

$$Gm \text{ commands} = \sum k \cdot \left(S_i, U\right). \tag{11}$$

5. Results and Analysis

The proposed FIM scheme is analyzed by developing a testbed using Open IMS core [28] to set up IMS entities including HSS, P-CSCFs, I-CSCF, and S-CSCF as illustrated in Figure 7. It is connected through IP network with four ARs where each one has a router and 802.11 g WLAN Access Point (AP). UE and CN are android phones that are connected to AN through WLAN AP. There are four ANs that individually contain SSID_1 to SSID_4 where UE disconnects from SSID_1 and reconnects to SSID_2. UE and APs establish session and exchange real-time transport protocol packets encoded with G.711 codec at 20 ms interval. UE also monitors the signaling strength of APs, and when the exponential smoothing value of the strength ($S_t = aS_{t-1} + (1 - a)a_t$, $0 \le a \le 1$) goes below the threshold value, then UE connects to new AP with strong signal strength, where a_t is signaling strength of AP at time t and S_t is the result at time t and $a = 0.5$ in the experiment as in [11]. For analysis of results, MDT of FIM is compared with MIP-IMS [2, 7] and standard IMS [17] scheme. Delay is considered as in [10], and number of hops and number of variations are considered as in [8]. Evaluation parameters for testbed are listed in Table 1. Moreover, for MIP-IMS, a router is deployed that maps the HoA with CoA and acts as HA. We have performed a number of handovers that can be categorized in three scenarios including Same P-CSCF (SP), Other Far P-CSCF (OFP), and Other Near P-CSCF (ONP) as discussed below.

(a) SP: It is handover scenario where UE disconnects from AN and then connects to a new AN but it gets attached to IMS with the same P-CSCF after handover.

(b) OFP: After handover to new AN, UE connects to a new P-CSCF that is physically located farther than old P-CSCF.

```
PCSCFold receives request
IF method == "HANDOVER" then
    Add IPSec SAs to T-IPSecSAs
    Path = discovered-PCSCF;
    Integrity-Protected = “yes”;
    Route to S-CSCF
ELSEIF method == ”REGISTER”
    route "REGISTER" to ICSCF
    IF status == "401"
        STATE Remove CK, IK
    ELSE
        STATE reply("500", "P-CSCF Error on Removing CK, IK");
    ENDIF
ENDIF
IF status == "200" then
    Route “OK” to UE
ENDIF
IF P-CSCFold != P-CSCFnew
    Delete IPSec SAs
ENDIF
```

Pseudocode 3: IPSec SAs handling at old P-CSCF.

Table 1: Evaluation parameters for testbed.

Testbed setup	
Parameters	Values
Network servers	P-CSCF, I-CSCF, S-CSCF, HSS
Servers' physical type	Wired physical
UEs' physical type	Wireless physical
Antenna type	Omni antenna
Parameter variations	
Delay	0–90 milliseconds
Number of hops	1–5
Number of handovers	1–10

(c) ONP: UE connects to new P-CSCF where the distance of new P-CSCF is physically less than the old P-CSCF from UE.

5.1. Media Disruption Time. It is the time delay that occurred due to credential transfer during handover. In standard IMS, handover MDT can be calculated as given in (12), where T_{UEtoARn} is the time taken for UE to connect to new AR and T_{UEtoP} is time taken by UE to get itself registered at P-CSCF. Time for REGISTER request to and from P-CSCF and S-CSCF is represented as T_{PtoS}. Time taken by REGISTER request to move between CSCF servers and HSS is denoted by $T_{\text{CSCFtoHSS}}$ and time for INVITE request for session establishment between UE and CN in SIP is represented as T_{UEtoCN}. Schemes [2, 4, 5] use Re-INVITE request after handover where three messages are exchanged between UE and CN. In MIP based schemes and FIM scheme, Re-INVITE method is considered for correct comparison. In standard IMS method, there is no scheme for handover so INVITE method is used. In our experiments, we implemented MIP based IMS mobility scheme using HA to measure MDT as given in (13), where $T_{\text{HA}} + T_{\text{UEtoCN}}$ represents the time taken for sending Binding Update to CN and HA. Similarly, FIM scheme is implemented to measure MDT as given in (14)-(15).

$$\begin{aligned}\text{MDT}_{\text{IMS}} &= 2T_{\text{UEtoARn}} + 4T_{\text{UEtoP}} + 4T_{\text{PtoS}} \\ &\quad + 6T_{\text{CSCFtoHSS}} + 8T_{\text{UEtoCN}}\end{aligned} \tag{12}$$

$$\begin{aligned}\text{MDT}_{\text{MIP}} &= 2T_{\text{UEtoARn}} + 2T_{\text{HA}} \\ &\quad + 5T_{\text{UEtoCN}} + 4T_{\text{UEtoP}} + 4T_{\text{PtoS}} \\ &\quad + 6T_{\text{CSCFtoHSS}}\end{aligned} \tag{13}$$

$$\begin{aligned}\text{MDT}_{\text{SP}} &= 2T_{\text{UEtoARn}} + T_{\text{UEtoPold}} + T_{\text{PoldtoS}} \\ &\quad + 3T_{\text{UEtoCN}} + T_{\text{PnewtoUE}}\end{aligned} \tag{14}$$

$$\begin{aligned}\text{MDT}_{\text{OFP\_ONP}} &= 2T_{\text{UEtoARn}} + T_{\text{UEtoPold}} + T_{\text{PoldtoS}} \\ &\quad + 2T_{\text{StoPnew}} + 3T_{\text{UEtoCN}} \\ &\quad + T_{\text{PnewtoUE}}.\end{aligned} \tag{15}$$

To evaluate MDT, we set $T_{\text{UEtoRAN}} = 10$ ms as in [29, 30] and $T_{\text{UEtoARo}} = 11$ ms, $T_{\text{ARotoARn}} = 5$ ms, $T_{\text{UEtoPold}} = 15$ ms, $T_{\text{PoldtoPnew}} = 7$ ms, and $T_{\text{PoldtoS}} = 10$ ms as in [10]. We considered Internet delay as 100 ms and $T_{\text{HA}} = 116$ ms, $T_{\text{UEtoCN}} = 128$ ms, and $T_{\text{HAtoCN}} = 114$ ms as in [10]. We assumed $T_{\text{UEtoARn}} = 10$ ms, $T_{\text{PnewtoUE}} = 16$ ms, and $T_{\text{StoPnew}} = 12$ ms. MDT versus delay between UE and new AR is obtained for SP, OFP, and ONP scenarios. It can be seen that, for IMS standard scheme, MDT is increasing with a high rate. In Figure 8(a) for SP scenario, for a delay of 30 ms, the MDT values are 338 ms, 408 ms, and 80 ms for IMS, MIP-IMS, and FIM, respectively,

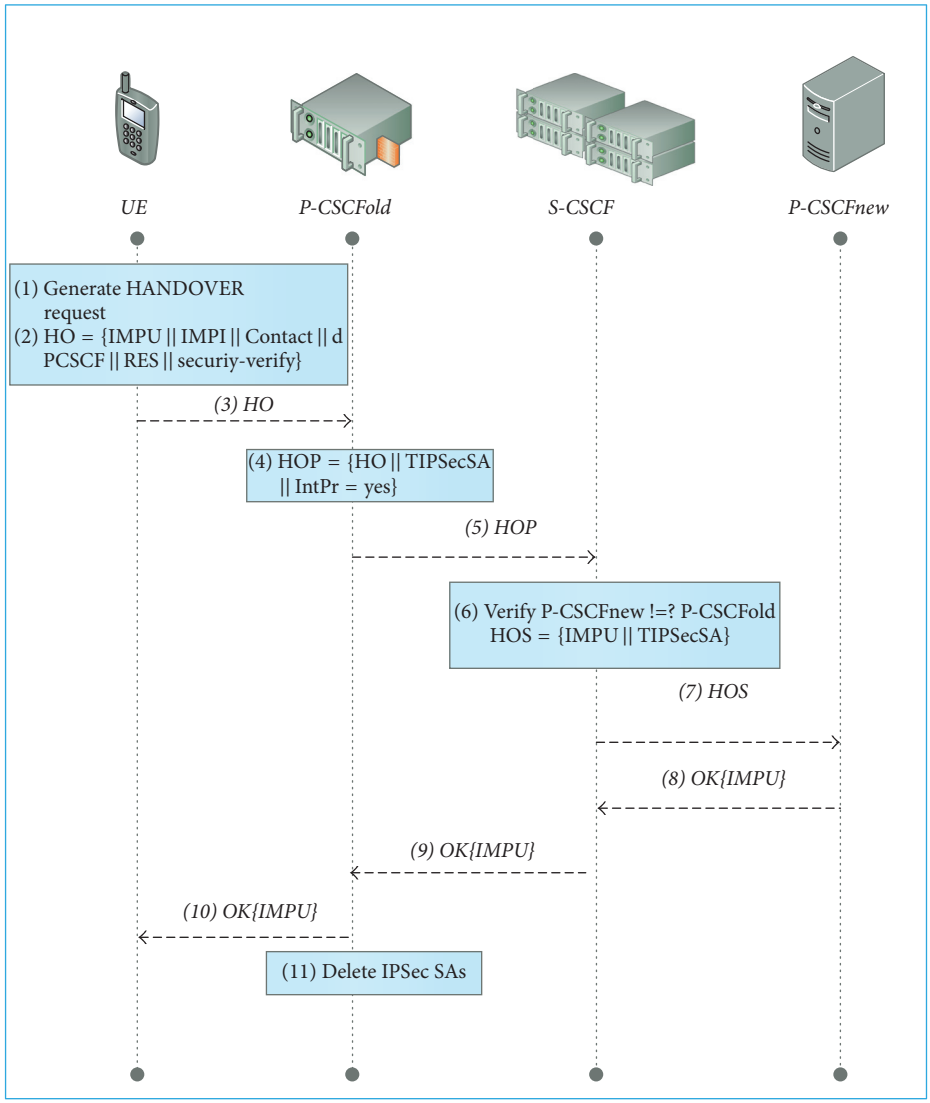


Figure 6: Proposed flow for FIM scheme.

where FIM requires the lowest MDT. It is also observed that when delay is increased three times, that is, 90 ms, then MDT values are 458 ms, 528 ms, and 122 ms for IMS, MIP-IMS, and FIM, respectively. In Figure 8(b) of OFP scenario, for a delay of 30 ms, the MDT values are 338 ms, 440 ms, and 169 ms for IMS, MIP-IMS, and FIM, respectively. It is also observed that when the delay is increased three times, that is, 90 ms, then MDT is 458 ms, 596 ms, and 229 ms for IMS, MIP-IMS, and FIM, respectively. FIM outperforms the others by consuming the lowest MDT.

In Figure 9, for ONP scenario, for a delay of 30 ms, the MDT values are 338 ms, 408 ms, and 100 ms for IMS, MIP-IMS, and FIM, respectively, where FIM outperforms the others by consuming the lowest disruption time.

Figure 10(a) elucidates that, in SP, for a delay = 40 ms, MDT values are 468 ms, 578 ms, and 110 ms for MIP-IMS, IMS, and FIM, respectively. Results prove the dominance of our proposed FIM scheme over preliminaries. It explores that existing schemes endure an abrupt change in MDT values as compared to a steady increase in proposed FIM scheme. In SP case, proposed scheme FIM reduces 71% compared to IMS scheme and it reduces 83% MDT compared to MIP-IMS scheme. Figure 10(b) elucidates that, in case of OFP, the MDT values at 40 ms delay are 468, 609, and 234 for IMS, MIP-IMS, and FIM methods, respectively. It explores that our proposed scheme is faster than the other two schemes. In case of OFP, FIM reduces 50% as compared to IMS scheme and it reduces 61% compared to MIP-IMS scheme. Figure 10(c) elucidates that, in case of ONP, for a delay of 40 milliseconds (ms) the MDT values are 468 ms, 578 ms, and 140 ms for MIP-IMS, IMS, and FIM scheme, respectively. Results are evident to elaborate the dominance of our proposed FIM scheme over preliminaries. In ONP case, proposed scheme FIM reduces 71% MDT compared to IMS scheme and it reduces 77% MDT compared to MIP-IMS scheme. Figure 10(d) shows the MDT versus delay between UE and new P-CSCF. It elucidates that

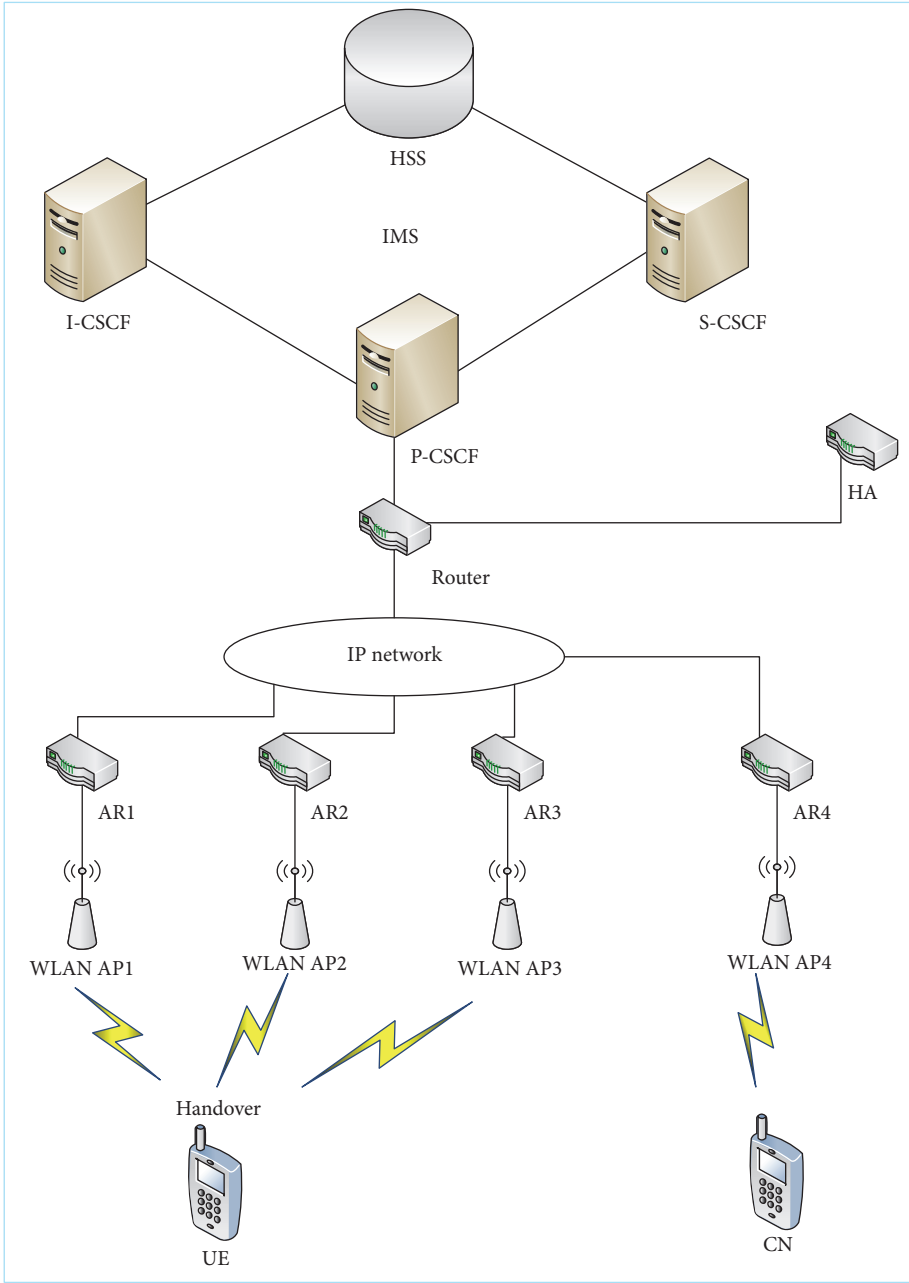


Figure 7: Experimental testbed scenario for FIM scheme.

for a delay of 24 ms the MDT values are 356 ms, 463 ms, and 178 ms for MIP-IMS, IMS, and FIM, respectively. In this case, proposed FIM scheme reduces 50% MDT compared to IMS scheme and it reduces 61% MDT compared to MIP-IMS scheme.

Figures 11(a)–11(c) elucidate handover scenario when the number of hops between UE and new AR is varied from 1 to 5. By presenting values in a sequence of (a) SP, (b) OFP, and (c) ONP for 3-hop scenario, the MDT values are 438 ms, 540 ms, and 567 ms for MIP-IMS and 364 ms, 378 ms, and 476 ms for IMS whereas our FIM scheme consumes only 119 ms, 189 ms, and 155 ms. Figures 11(d)–11(f) elucidate the scenario when the number of hops is varied between UE and CN. By following the same sequence as above, the MDT values are 497 ms, 678 ms, and 418 ms for MIP-IMS and 598 ms, 501 ms, and 544 ms for IMS, whereas FIM consumes only 140 ms, 287 ms, and 121 ms for SP, OFP, and ONP, respectively. These results are evident to elaborate the dominance of our proposed FIM scheme over preliminaries. Table 2 shows the comparison of the improvement of proposed scheme with IMS and MIP-IMS schemes in terms of percentage for numerical and testbed. It proves that FIM is better in reducing MDT than others when number of hops is varied between UE and new AR and between UE and CN as well. It is observed

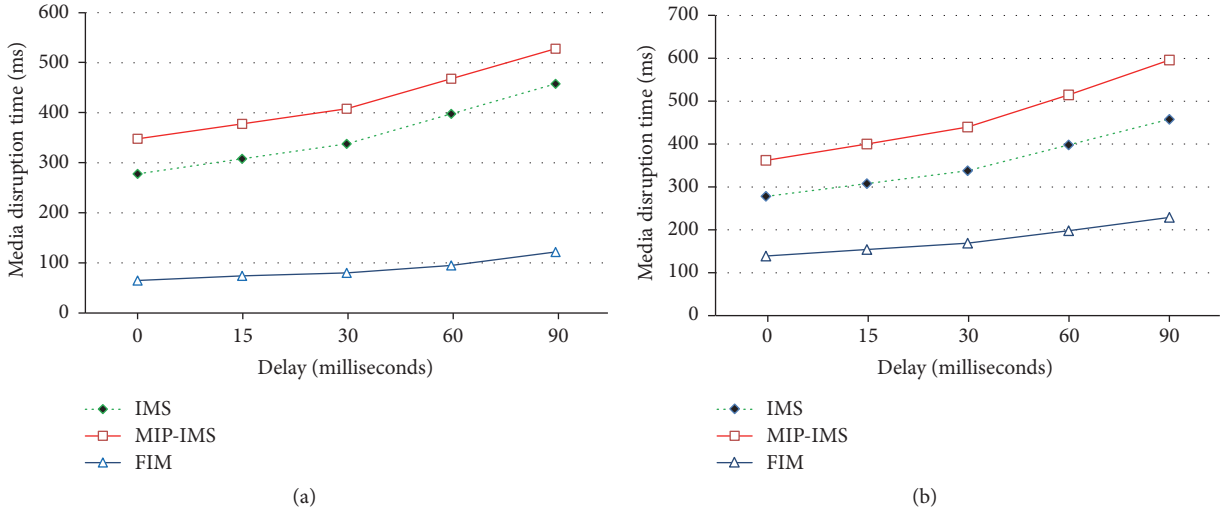


Figure 8: MDT versus delay between UE and AR in (a) for SP and in (b) for OFP.

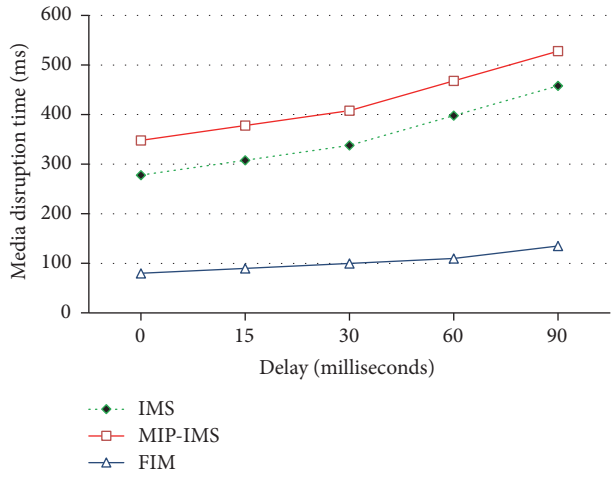


Figure 9: Media disruption time versus delay between UE and ARn in ONP.

that, in case of ONP, MDT is lesser than OFP and it is further reduced in case of SP.

5.2. Packet Loss. We have set $G = 50$ pkts/sec as used by [31]. On a constant handover delay, we have measured loss of packets for IMS, MIP-IMS, and FIM. Packet loss is directly proportional to delay of handover [26] as explored in Figure 12(b). During handover, the loss of packets is 44400 bytes, 57600 bytes, and 22200 bytes for IMS, MIP-IMS, and FIM schemes, respectively. FIM reduces the packet loss by 50% compared to IMS and 61% compared to MIP-IMS.

5.3. Number of Commands. Our solution reduces the number of *Gm* commands between UE and P-CSCF and *Mw* commands between P-CSCF and S-CSCF. Moreover it totally eliminates *Cx* commands between CSCSF servers and HSS. We compare the number of messages of standard IMS handoff scenario and MIP handoff scenarios with FIM handoff approach. Table 2 shows the number of commands for IMS handoff, MIP-IMS handoff, and FIM handoff schemes. Figure 12(a) elucidates that the values of *Gm*, *Mw*, and *Cx* are 60, 60, and 60, respectively, for 10 handovers in case of IMS and MIP-IMS methods. In comparison, values for our proposed scheme are significantly low, that is, 20, 20, and 0, respectively. Our scheme completely removes the need for *Cx* commands.

5.4. SIP Session Delay. To test the delay of our proposed scheme (FIM) for SIP messages, an experiment is run after handover. The delay time of SIP session is captured by using Wireshark. Figure 13 shows the time of SIP session that confirms the reduction of delay.

6. Conclusion

NGMN provide a variety of multimedia applications with huge data of live video streaming from cameras or recorded videos for delivering to users. During such data transmissions, mobile UE has to perform handover that can lead to

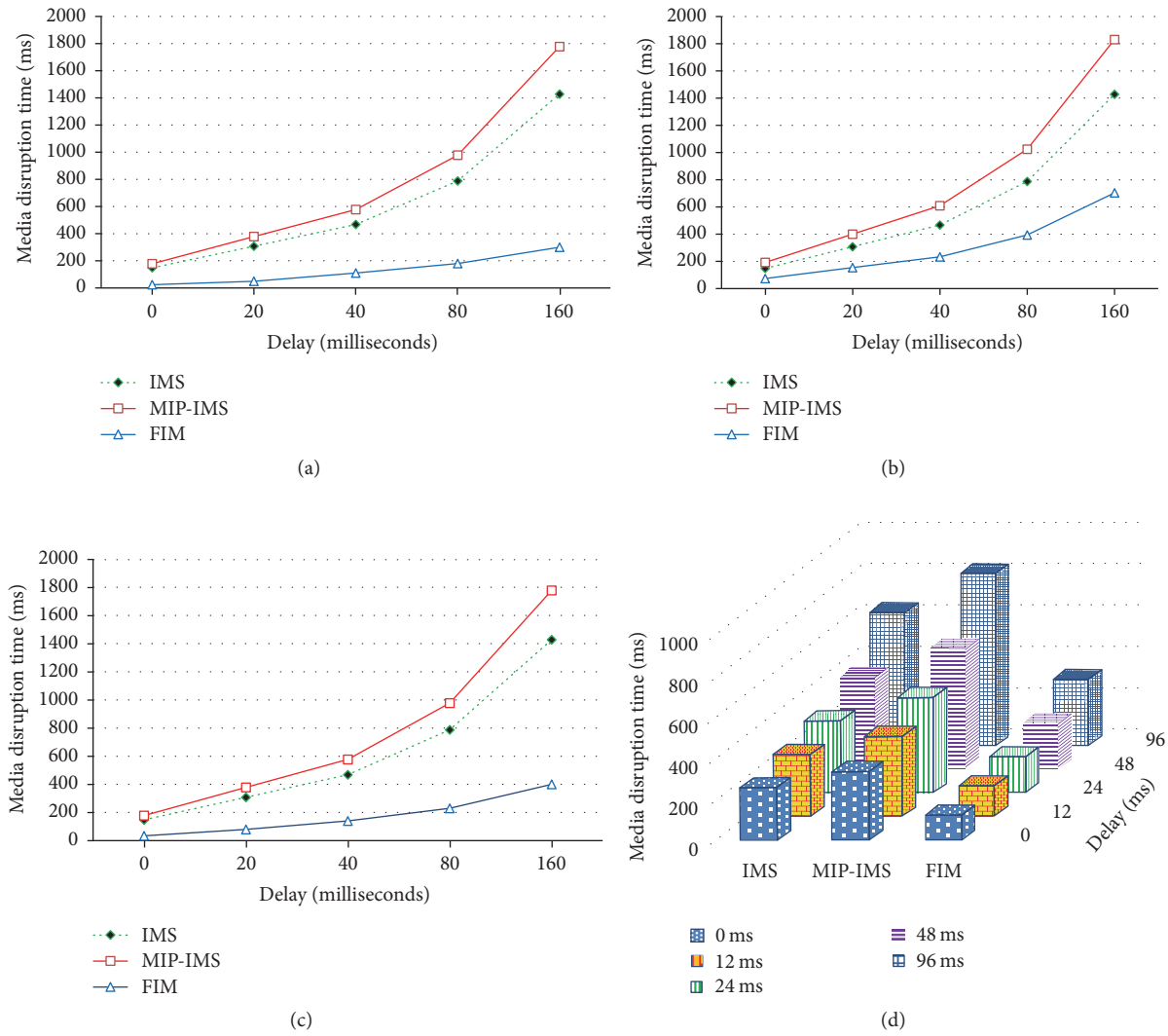


Figure 10: MDT versus delay between UE and CN for (a) SP, (b) OFP, (c) ONP, and (d) MDT versus delay between UE and new P-CSCF.

Table 2: Improvement of FIM scheme over other schemes.

Number of hops	Scenarios →	SP		OFP		ONP	
	Schemes	Numerical	Testbed	Numerical	Testbed	Numerical	Testbed
Between UE and AR	IMS	76%	67%	50%	50%	71%	65%
	MIP-IMS	81%	72%	61%	65%	77%	71%
Between UE and CN	IMS	71%	71%	50%	42%	71%	71%
	MIP-IMS	83%	76%	61%	57%	77%	77%

latency and overhead due to disconnection and reregistration. We have presented a session aware registration scheme named FIM for handover scenario where session credentials are transferred without a call session interruption or disconnection. We have developed a testbed using Open IMS to perform experiments by configuring different evaluation parameters for IMS servers and UE. Moreover, algorithms are also proposed for authentication and transfer of IPSec SAs during handover scenario. We have also measured the SIP session delay during handover by using Wireshark. Furthermore the numerical and testbed values are also compared. Results elucidate that FIM reduces MDT by 50% and 60% as compared to IMS and MIP-IMS in OFP handover scenario. In case of SP and ONP, our scheme achieves an average of 70% less MDT consumption. FIM reduces 67% number of *Gm* commands, 67% of *Mw* commands, and 100% of

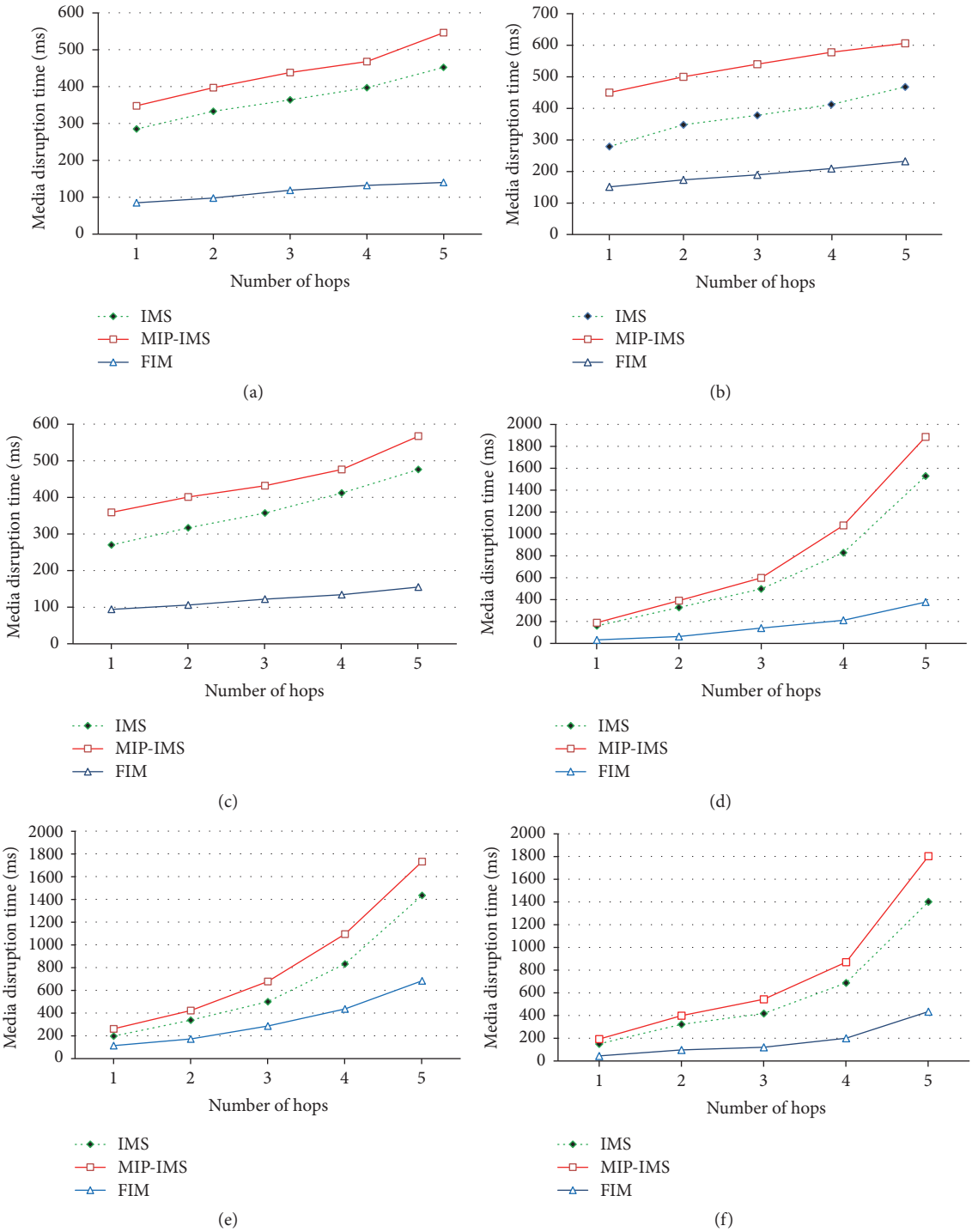


Figure 11: MDT versus number of hops between UE and new AR is presented for (a) SP, (b) OFP, and (c) ONP and between UE and CN is presented in (d) SP, (e) OFP, and (f) ONP scenarios.

Cw commands. It also reduces 50% and 61% packet loss as compared to IMS and MIP-IMS, respectively. Results prove the dominance of FIM scheme over preliminaries.

Notations

UE: User equipment
AN: Access network
QoS: Quality of service
MDT: Media disruption time
AR: Access Router
IPSec SAs: IPSec Security Associations
Mw: Commands between CSCFs
Gm: Command from UE to IMS

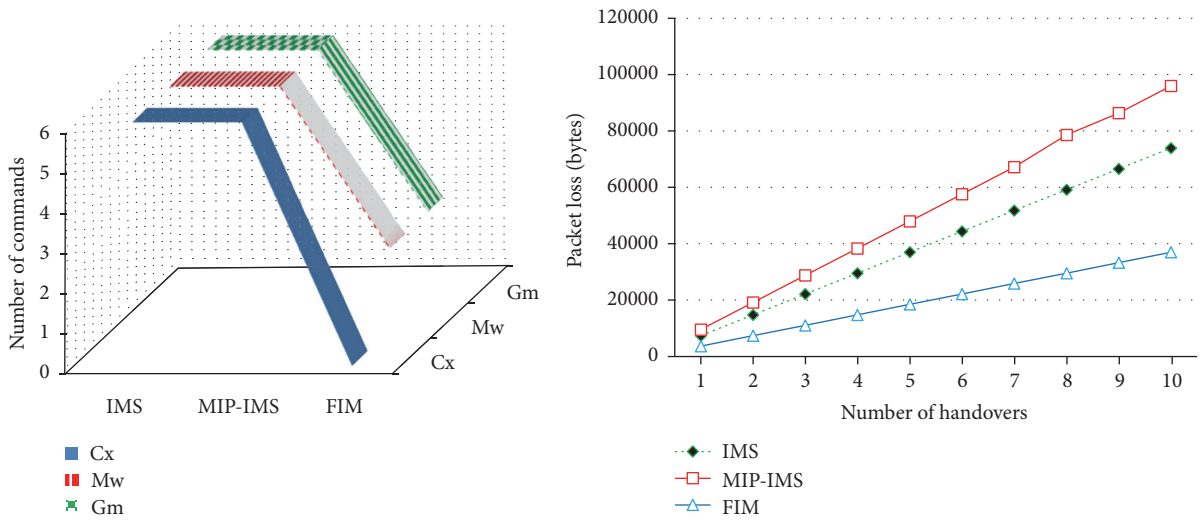


Figure 12: The number of *Cx*, *Mw*, and *Gm* commands is presented in (a) and packet loss is presented in (b).

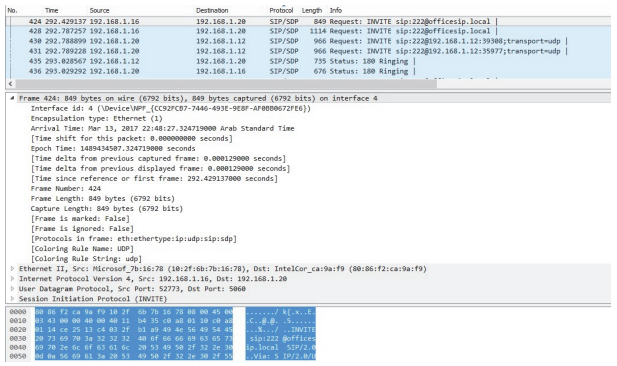


Figure 13: Analysis of FIM using Wireshark.

Cx:	Command from CSCFs to HSS
BS:	Base Station
T:	Time
AP:	Access Point
HA:	Home Agent
CoA:	Care of Address
HoA:	Home Address
BU:	Binding Update
BU-Ack:	BU Acknowledgment
RES:	Calculated response
IntPr:	Integrity-Protected
dPCSCF:	Discovered P-CSCF
TIPSecSA:	Old IPSec SA.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors extend their appreciation to the International Scientific Partnership Program ISPP at King Saud University for funding this research work through ISPP no. 0033.

References

[1] M. Abu-Lebdeh, J. Sahoo, R. Glitho, and C. W. Tchouati, "Cloudifying the 3GPP IP multimedia subsystem for 4G and beyond: a survey," *IEEE Communications Magazine*, vol. 54, no. 1, pp. 91–97, 2016.

[2] N. Akkari, "An IMS-based integration architecture for WiMax/LTE handover," *Computer Networks*, vol. 57, no. 18, pp. 3790–3798, 2013.

[3] B. Moon, "Fast and secure session mobility in IMS-based vertical handover scenario," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 9, pp. 171–188, 2014.

[4] E. P. Edward, "A novel seamless handover scheme for WiMAX/LTE heterogeneous networks," *Arabian Journal for Science and Engineering*, vol. 41, no. 3, pp. 1129–1143, 2016.

[5] E. Prince Edward and V. Sumathy, "Performance analysis of a context aware cross layer scheme for fast handoff in IMS based integrated WiFi-WiMax networks," *Pervasive and Mobile Computing*, vol. 17, pp. 79–101, 2015.

[6] T. Renier, K. L. Larsen, G. Castro, and H.-P. Schwefel, "Mid-session macro-mobility in IMS-based networks," *IEEE Vehicular Technology Magazine*, vol. 2, no. 1, pp. 20–27, 2007.

[7] K. L. Larsen, E. V. Matthiesen, H.-P. Schwefel, and G. Kuhn, "Optimized macro mobility within the 3GPP IP multimedia subsystem," in *Proceedings of the 2nd International Conference on Wireless and Mobile Communications (ICWMC '06)*, Bucharest, Romania, July 2006.

[8] A. Bagubali, V. Prithiviraj, and P. S. Mallick, "Performance analysis of IMS based LTE and WIMAX integration architectures," *Alexandria Engineering Journal*, vol. 55, no. 4, pp. 3229–3238, 2016.

[9] J. T. Renier, F. Hanane, and S. Hans-Peter, "MIP-based enhanced mid-session macro-mobility for IMS-controlled state-full applications," in *Proceedings of the 10th International Symposium on Wireless Personal Multimedia Communications*, pp. 671–675, Jaipur, India, 2007.

[10] R. Farahbakhsh and N. Movahhedinia, "Seamless handover for IMS over mobile-IPv6 using context transfer," 2012, https://arxiv.org/abs/1208.1207.

[11] M. Ito, S. Komorita, Y. Kitatsuji, and H. Yokota, "IMS-based fast session handover with available network resources discovery of access networks," *Journal of Information Processing*, vol. 20, no. 1, pp. 308–318, 2012.

[12] A. Nazari, J. But, P. Branch, and H. Vu, "PRIME: Pre-registration for IMS mobility enhancement," in *Proceedings of the 13th IEEE International Conference on Multimedia and Expo (ICME '12)*, pp. 920–924, Melbourne, VIC, Australia, July 2012.

[13] A. Nazari, P. Branch, J. But, and H. L. Vu, "UPTIME: an IMS-based mobility framework for next generation mobile networks," *Wireless Networks*, vol. 20, no. 7, pp. 1967–1979, 2014.

[14] J. Rosenberg, S. Henning, C. Gonzalo et al., "SIP: session initiation protocol," RFC 3261, 2002.

[15] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," RFC 3775, IETF, 2004.

[16] O. Khattab and O. Alani, "A survey on media independent handover (MIH) and IP multimedia subsystem (IMS) in heterogeneous wireless networks," *International Journal of Wireless Information Networks*, vol. 20, no. 3, pp. 215–228, 2013.

[17] M. Poikselkä and M. Georg, *The IMS: IP Multimedia Concepts and Services*, John Wiley & Sons, Hoboken, NJ, USA, 2013.

[18] R. Koodli, "Mobile IPv6 Fast Handovers," RFC Editor RFC5568, 2009.

[19] N. H. Thanh, L. T. Hang, N. Q. Thu, V. V. Yem, and N. X. Dung, "Multimedia session continuity with context-aware capability in IMS-based network," in *Proceedings of the 6th International Symposium on Wireless Communication Systems (ISWCS '09)*, pp. 383–387, Tuscany, Italy, September 2009.

[20] A. De La Oliva, A. Banchs, I. Soto, T. Melia, and A. Vidal, "An overview of IEEE 802.21: media-independent handover services," *IEEE Wireless Communications*, vol. 15, no. 4, pp. 96–103, 2008.

[21] S.-R. Yang and W.-T. Chen, "SIP multicast-based mobile quality-of-service support over heterogeneous IP multimedia subsystems," *IEEE Transactions on Mobile Computing*, vol. 7, no. 11, pp. 1297–1310, 2008.

[22] X. Yang and A. Agarwal, "Multicast mobility in SIP layer," in *Proceedings of the IEEE 59th Vehicular Technology Conference, VTC2004-Spring: Towards a Global Wireless World*, Milan, Italy, May 2004.

[23] P. Pongpaibool, P. Sotthivirat, S. I. Kitisin, and C. Srisathapornphat, "Fast duplicate address detection for mobile IPv6," in *Proceedings of the 15th IEEE International Conference on Networks (ICON '07)*, pp. 224–229, Adelaide, SA, Australia, aus, November 2007.

[24] X. Yan, Y. A. Şekercioğlu, and S. Narayanan, "A survey of vertical handover decision algorithms in Fourth Generation heterogeneous wireless networks," *Computer Networks*, vol. 54, no. 11, pp. 1848–1863, 2010.

[25] P. Bellavista, A. Corradi, and L. Foschini, "An IMS vertical handoff solution to dynamically adapt mobile multimedia services," in *Proceedings of the 13th IEEE Symposium on Computers and Communications (ISCC '08)*, pp. 764–771, Marrakech, Morocco, July 2008.

[26] K. S. Munasinghe and A. Jamalipour, "Route optimization for roaming heterogeneous multi-homed mobile networks," in *Proceedings of the 4th International Conference on Signal Processing and Communication Systems (ICSPCS '10)*, Gold Coast, QLD, Australia, December 2010.

[27] G. Camarillo and M.-A. García-Martín, *The 3G IP Multimedia Subsystem (IMS): Merging The Internet and The Cellular Worlds*, John Wiley & Sons, Hoboken, NJ, USA, 2007.

[28] T. Magedanz, D. Witaszek, and K. Knuettel, "The IMS playground @ FOKUS-an open testbed for next generation network multimedia services," in *Proceedings of the 1st International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities (Tridentcom '05)*, pp. 2–11, Trento, Italy, February 2005.

[29] J. Kempf, "Problem description: reasons for performing context transfers between nodes in an IP access network," RFC Editor RFC3374, 2002.

[30] X. Pérez-Costa, M. Torrent-Moreno, and H. Hartenstein, "A performance comparison of Mobile IPv6, Hierarchical Mobile IPv6, fast handovers for Mobile IPv6 and their combination," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, no. 4, pp. 5–19, 2003.

[31] L. Cai, Y. Xiao, X. Shen, L. Cai, and J. W. Mark, "VoIP over WLAN: Voice capacity, admission control, QoS, and MAC," *International Journal of Communication Systems*, vol. 19, no. 4, pp. 491–508, 2006.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 5790161, 12 pages
https://doi.org/10.1155/2017/5790161

Research Article

Energy-Efficient Unequal Chain Length Clustering for Wireless Sensor Networks in Smart Cities

Mohammad Baniata and Jiman Hong

School of Computer Science & Engineering, Soongsil University, Seoul, Republic of Korea

Correspondence should be addressed to Jiman Hong; gmaneos@gmail.com

Received 25 April 2017; Accepted 1 August 2017; Published 7 September 2017

Academic Editor: Syed Hassan Ahmed

The recent advances in sensing and communication technologies such as wireless sensor networks (WSN) have enabled low-priced distributed monitoring systems that are the foundation of smart cities. These advances are also helping to monitor smart cities and making our living environments workable. However, sensor nodes are constrained in energy supply if they have no constant power supply. Moreover, communication links can be easily failed because of unequal node energy depletion. The energy constraints and link failures affect the performance and quality of the sensor network. Therefore, designing a routing protocol that minimizes energy consumption and maximizes the network lifetime should be considered in the design of the routing protocol for WSN. In this paper, we propose an Energy-Efficient Unequal Chain Length Clustering (EEUCLC) protocol which has a suboptimal multihop routing algorithm to reduce the burden on the cluster head and a probability-based cluster head selection algorithm to prolong the network lifetime. Simulation results show that the EEUCLC mechanism enhanced the energy balance and prolonged the network lifetime compared to other related protocols.

1. Introduction

As cities come to be more and more crowded, public safety, efficient transport, preserving energy, lessening expenses, and improving the quality of life are becoming indispensable. Therefore, the smart city term is gaining popularity among scientific researchers. Chen [1] defined smart cities as "taking the advantage of communications and sensor capabilities sewn into the cities infrastructures to optimize electrical, transportation, and other logistical operations supporting daily life, thereby improving the quality of life for everyone."

With recent advances in microelectromechanical systems technology and the low price of manufacturing very small low powered sensing devices and deploying them on a large scale, smart cities are becoming technically and economically possible. Wireless sensor networks (WSNs) are equipped with computing, sensing, and wireless communication capabilities. Moreover, due to their small size and low cost, sensors are normally deployed in different environments, where they can collaborate among themselves to form a special sensor network. Since WSN may consist of tens, hundreds, or thousands of sensors, it helps to achieve fault tolerant, reliable, and accurate sensor network covering a wider region [2].

Given these capabilities, WSN have recently stood out as a platform on which to deploy different kinds of important smart city applications [3, 4], such as power systems [5, 6], transportation and roadside applications [7, 8], healthcare [9, 10], and air pollution monitoring [11]. Figure 1 presents the various kinds of application using WSN deployment in smart cities.

However, wireless sensor nodes have limitations in their capabilities, such as small memory, limited energy, and low processing power [12, 13]. Sensors consume energy in sensing, data processing, and communicating without a permanent power supply along with the difficulty of replacing them. Communication is often the most energy-intensive. As a result, it is generally accepted that a WSN expires when its battery runs out, which affects its performance, quality, and lifetime.

Therefore, the lifetime of a network in critical applications such as power systems is decisive because the ability to save human lives as long as possible is very urgent. Due to the strict energy constraints, prolonging the lifetime of the sensor

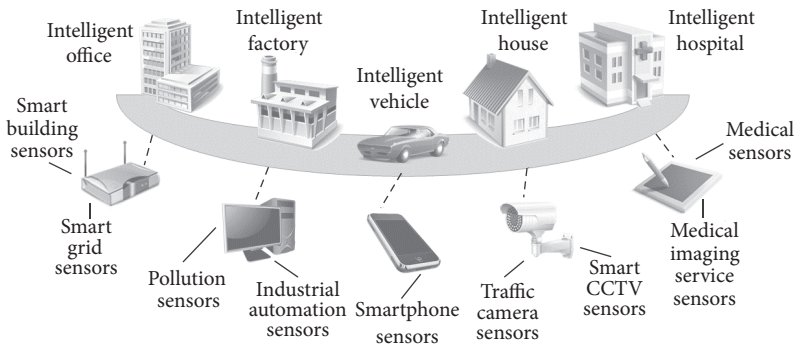


Figure 1: Various applications of WSN in a smart city.

networks and reducing energy consumption are the most critical problem that must be considered when designing a WSN for a smart city.

If each and every sensor starts to communicate their data directly to the base station, the energy of the sensor nodes will be consumed quickly due to the long distances to the base station. In addition, congestion and collisions will consume more energy due to data retransmission. Clustering mechanism for the WSN is the approach used for solving these concerns. In clustering mechanism, nodes within a cluster send their data over shorter distances to their respective cluster heads, which aggregate sensed data from the surrounding general nodes into small sets of meaningful information and pass it on to the base station. Furthermore, aggregating and fusing data at the cluster head within a cluster helps to enucleate data redundancy.

Being a cluster head node means that it consumes more energy than general nodes, but this burden can be alleviated by selecting a few sensor nodes as a cluster head and rotating this role among the cluster members. The cluster heads can send data directly or use multihop to send data to the base station. Many protocols have been proposed to improve their energy consumption and the network lifetime, such as the Low-Energy Adaptive Clustering Hierarchy (LEACH) [14], Power-Efficient Gathering in Sensor Information Systems (PEGASIS) [15], Hybrid Energy-Efficient Distributed Clustering (HEED) [16], Energy-Efficient Unequal Clustering (EEUC) [17], and the Fast Local Clustering Service (FLOC) [18]. However, the unsolved issue of energy consumption and hotspot [19] in the previous clustering protocols for WSN still exists.

In order to overcome several hotspot issues, we present and evaluate the EEUCLC, Energy-Efficient Unequal Chain Length Clustering protocol for the systemic gathering of data from the sensing environments where each node for every round has an information packet that should be transmitted to the base station. EEUCLC is a distributed competitive protocol where cluster heads are selected by localized competition. EEUCLC groups the networks into several unequal intracluster chain lengths and finds suboptimal multihop routes over the cluster heads to propagate the aggregation data to the base station. In EEUCLS protocol, the node with the highest residual energy and shortest distance to base station has the best chance to become the cluster head. Therefore the nodes transmission radius decreases as its distance to base station decreases, which leads to creating smaller cluster sizes with shorter sensors chain length than those farther away from the base station. Hence, the accelerated consumed energy due to intracluster data processing and long distances to respective cluster heads will be reduced. EEUCLC also considers distance and residual energy to find suboptimal paths between cluster head nodes to the base station and uses these paths randomly for every data transmission occasion.

2. Related Works

Many clustering protocols [14–16, 18] and unequal clustering protocols [20–24] have been suggested to minimize the energy consumption of WSNs.

LEACH presented in [14] is a well-known clustering protocol for wireless sensor networks. LEACH is a distributed cluster formation without centralized control where each cluster head gathers data from all of the nodes to reduce global communication. LEACH rotates the cluster head role among all the nodes to distribute the energy load evenly. Also, an improved scheme of LEACH was proposed, named LEACH-C [25]. In LEACH-C, a centralized algorithm at the base station makes cluster formation.

In [15], Lindsey and Raghavendra proposed PEGASIS. PEGASIS is an enhancement of the LEACH protocol and makes a communication chain using a TSP (Traveling Sales Person) heuristic. Each node only communicates with two close neighbors along the communication chain. Only one node is chosen to send the data to the base station instead of multiple nodes. All nodes in the chain will take turns sending to the base station.

HEED presented in [16] is a multihop clustering protocol. It extends LEACH by considering range limits of the wireless communication and intracluster communication cost. It is considered as a distributed cluster formation where each node makes a decision based on local information. The probability for each node to become a tentative cluster head depends on its residual energy, and all the tentative heads are competing for becoming the final cluster heads. The final cluster heads are selected according to the intracluster communication cost.

In [17], Li et al. proposed EEUC, which is an Energy-Efficient Unequal Clustering protocol. It divides the nodes

into clusters of unequal size. The clusters closer to the base station have smaller sizes than those farther away from the base station. Thus cluster heads closer to the base station can preserve energy for the intercluster data forwarding.

FLOC [18] is a distributed clustering technique that uses the solid-disc property to minimize overlap between multiple clusters. A node communicates reliably with others within its unite distance defined as the inner-band (I-band) range and communicates unreliably (only a percentage of messages go through) with nodes within $m \leq 2$ distance defined as the outer band (O-band) range. FLOC is fast and scalable and achieves clustering in O(1) time.

In [20], Ever et al. proposed UHEED which is an unequal clustering algorithm that tries to reduce energy consumption and avoid the early death of cluster heads closer to the base station by adjusting the completion radius of the HEED protocol based on the equation being used in EEUC.

In [21], Ren et al. proposed the Unequal Clustering Scheme based on LEACH for better energy balancing and extending network lifetime. The energy ratio and competition distance are used as two elements to join the cluster head selection. In this scheme, nodes can adjust their transmission power based on receiving a distance matrix from the base station.

EADUC [22], Energy-Aware Distributed Unequal Clustering, protocol is a distributed unequal clustering technique that supports both homogeneous and heterogeneous networks. The clustering formation is divided into three subphases: neighbor node information collection phase, cluster head competition phase, and cluster formation phase. Cluster heads with higher residual energy and being further away from the base station will have a larger cluster size.

UCA [23], location-based unequal clustering algorithm, is another unequal clustering algorithm where each cluster head has a different cluster size based on its distance to the base station. LUCA forms a cluster which has a smaller cluster size as the distance to the base station decreases.

In [24], Yang and Zhang proposed EB-UCP, Energy-Balancing Unequal Clustering Protocol, where the closer nodes to the base station have a higher probability of being selected as cluster heads. In EB-UCP, the network is split into layers, and the nodes in each layer have their own probability of cluster head selection.

3. Problem Statement

3.1. Accelerated Energy Consumption Problem. In a large-scale network, the relationship between network size and cluster size is proportional where the number of nodes in the cluster increases as the network size increases. It means that cluster heads bear a heavy burden and consume a considerable amount of energy. Using the single hop forwarding model within the cluster, where each node sends its data directly to their respective cluster head, leads to an increase in the number of transmitted messages to the cluster head from nodes within the cluster.

Congestion and collisions can also take place, which in turn will consume more energy due to data retransmission. Other problems arise if the cluster head is far away from the nodes. In this case, some nodes will require a large amount of transmission power to convey their data; thus the energy of the nodes will be drained quickly.

Using multihop communication models between several cluster heads to forward the aggregated data to the base station is considered as a good solution for cluster heads that may not be able to communicate directly with the base station. However, using this method might be considered as another factor causing high energy consumption by the cluster head.

It happens in two different ways. First, if a cluster head is nearest to the base station it will act as a router for other nodes and is burdened with heavy relay traffic. Second, the best single route between a cluster head node to the base station will be found based on cluster head residual energy and the energy of the transmitter. This route is used for every communication. This will drain the energy of the cluster heads along this path. Therefore, the fundamental goal is to overcome these hot spot problems.

The protocol considers the periodic selection of cluster heads based on the nodes residual energy and distance to the base station. Furthermore, we present unequal chain clustering and a suboptimal multihop routing protocol to alleviate the burden on the cluster heads and prolong the lifetime of a sensor network.

3.2. Network and Radio Models. We consider a WSN of hundreds or thousands of N energy-constrained sensor nodes, which are uniformly deployed over a sensing environment. We denote node i by s_i and the identical sensor node set $S = \{s_1, s_2, \ldots, s_n\}$, where $1 \leq i \leq |S|$ and $|S| = N$. In this work, our model sensor network is based on the following assumptions:

(i) The base station is fixed and situated at far distance from the sensor nodes.

(ii) Links between the nodes are symmetric.

(iii) Nodes in the network are homogeneous.

(iv) Energy is constrained.

(v) The network is location aware, that is, equipped with a GPS capable antenna.

(vi) The nodes are able to utilize power control when transmitting data.

(vii) The nodes are not mobile.

We adapt the same radio model described in Heinzelman et al. [26]. In this radio model, a radio dissipated energy to run the transmitter and transmitter amplifier. In addition, the radio dissipated energy to run the receiver circuitry. Both the free space model (d^2 power loss) and the multipath fading (d^4 power loss) channel models are used, where the dissipation energy to run the transmitter amplifier depends on the distance between the transmitter and receiver. In this radio model, the energy dissipated to transmit 1-bit message for a distance d is computed as follows.

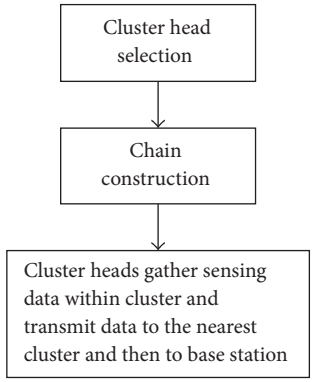


Figure 2: Flowchart of EEUCLC.

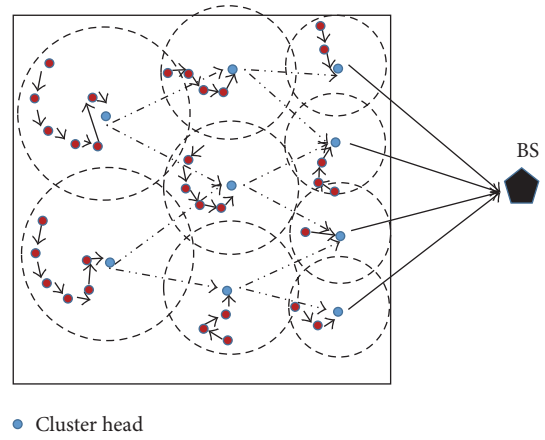


Figure 3: An overview of the EEUCLC technique (BS = base station).

Transmitting

$$E_{\mathrm{TX}}(\mathtt{l},\mathtt{d}) = E_{\mathrm{TX-elec}}(\mathtt{l}) + E_{\mathrm{TX-amp}}(\mathtt{l},\mathtt{d}) = \begin{cases} lE_{\mathrm{elec}} + E\epsilon_{\mathrm{fs}}d^2, & \mathtt{d} < d_0, \\ lE_{\mathrm{elec}} + E\epsilon_{\mathrm{mp}}d^4, & \mathtt{d} \geq d_0. \end{cases} \quad (1)$$

Receiving

$$E_{\mathrm{RX}} = E_{\mathrm{RX-elec}}(\mathtt{l}) = lE_{\mathrm{elec}}. \quad (2)$$

4. Proposed Algorithm

Figure 2 shows the flow chart and Figure 3 shows a general overview of the EEUCLC protocol, respectively. The EEUCLC protocol consists of three phases: (1) cluster head selection, (2) chain formation, and (3) data collection and transmission. In Figure 3, the circles of unequal size represent clusters of unequal chain length and the routes among the cluster heads clarify the multihop forwarding model.

4.1. Cluster Head Selection Phase. When clusters are being created, several candidate cluster heads are selected to compete to become final cluster heads in the current round. Periodic reselection within the cluster is performed to balance the energy consumption of each cluster.

For candidate cluster head selection, EEUCLC makes a tradeoff between the nodes residual energy and distance to the base station. A node having low residual energy with a long distance to the base station will not be selected to compete for the final cluster head. Before a node starts carrying out EEUCLC, it sets its probability to become a candidate cluster head, $\mathrm{CCH}_{\mathrm{prob}}$, as follows:

$$\mathrm{CCH}_{\mathrm{prob}} = a \cdot \frac{(d_{\max} - d(s_i, \mathrm{BS}))}{(d_{\max} - d_{\min})} + (1-a) \cdot \frac{E_{\mathrm{residual}}}{E_{\max}}, \quad (3)$$

where E_{residual} denotes the estimated current residual energy; $E_{\max}$ is the maximum energy, which is congruous for all nodes; $d_{\max}$ and $d_{\min}$ denote the maximum and minimum distance between sensor nodes and the base station, respectively. $d(s_i, \mathrm{BS})$ is the distance between s_i node and the base station, and a is a constant coefficient.

For a node to become a candidate cluster head, the value of $\mathrm{CCH}_{\mathrm{prob}}$ should be larger or equal to a certain threshold G. Nodes below this certain threshold value keep sleeping until the cluster head selection phase terminates. The purpose of using $\mathrm{CCH}_{\mathrm{prob}}$ is to limit the initial cluster head announcement and prevent nodes far away from the base station with low residual energy from consuming more energy due to an additional packet transmission overhead while competing to become the final cluster head.

After the candidate cluster heads have been selected, competition starts for the final cluster head phase. A candidate cluster head only becomes a final cluster head if it has greater residual energy than the other nodes in its competition radius. In order to create unequal intracluster chain lengths, we need to control the cluster size. In EEUCLC, intracluster chains closer to the base station have a shorter length than those farther away.

To achieve our objective, the cluster head competition radius should decrease as its distance to the base station decreases. To compute the cluster head competition radius, we use the same equation being used in EEUC:

$$s_i.R_{\mathrm{comp}} = 1 - c\frac{(d_{\max} - d(s_i, \mathrm{BS}))}{(d_{\max} - d_{\min})} R^0_{\mathrm{comp}}, \quad (4)$$

where c is a constant between 0 and 1, and R^0_{comp} is the initial cluster range used in remote areas from the base station.

Figure 4 shows an uneven competition radius of candidate cluster heads where candidate cluster heads S_3 and S_4 cannot both be final cluster heads because they are in the competition radius of each other, while the opposite occurs for S_1 and S_2.

Competition for final cluster head selection messages is broadcast with R^0_{comp}. In lines 5–29 of Algorithm 1, competition for final cluster heads starts when each candidate cluster head broadcasts a *CompeteChainHead_MSG* containing residual energy and competition radius and sets a random wait time.

If no *CompeteChainHead_MSG* is received within a random wait time, the candidate head node broadcasts

(1) **if** $\mathrm{CCH}_{\mathrm{prob}} \geq G$ **then**
(2) $beCandidateHead \leftarrow True$
(3) **end if**
(4) **if** beCandidateHead = True **then**
(5) $CompeteChainHead\_Msg(ID, R_{\mathrm{comp}}, E_{\mathrm{residual}})$
(6) **else**
(7) **EXIT**;
(8) **end if**
(9) On receiving a *CompeteChainHead_MSG* form node s_j
(10) **if** $d(s_i, s_j) < s_j.R_{\mathrm{comp}}$ OR $d(s_i, s_j) < s_i.R_{\mathrm{comp}}$ **then**
(11) Add s_j to s_i cluster group
(12) **end if**
(13) **while** beCandidateHead = True **do**
(14) **if** $s_i.E_{\mathrm{residual}} > s_j.E_{\mathrm{residual}}, \forall s_j \in s_i.Clustergroup$ **then**
(15) $FinalChainHead\_Msg(ID)$
(16) Exit
(17) **end if**
(18) On receiving *FinalChainHead_MSG* form node s_j
(19) **if** $s_j \in s_i.Clustergroup$ **then**
(20) *Resignation_MSG*(*ID*) and then EXIT
(21) **end if**
(22) On receiving a *Resignation_MSG* form node s_j
(23) **if** $s_j \in s_i.Clustergroup$ **then**
(24) Remove s_j from s_i cluster group
(25) **end if**
(26) **end while**
(27) *ChainHead_ADV_MSG*(*ID*).*Send*()
(28) On not receiving a *CompeteChainHead_MSG* form node s_j
(29) *ChainHead_ADV_MSG*(*ID*).*Send*()

Algorithm 1: Cluster head selection.

ChainHead_ADV_MSG containing *HeadNode_id*. In the case where the candidate head node receives *CompeteChainHead_MSG* from another candidates node, if $d(s_i, s_j) < s_j.R_{\mathrm{comp}}$ or $d(s_i, s_j) < s_i.R_{\mathrm{comp}}$, then they become part of one cluster group.

Each candidate cluster head in the cluster group checks the residual energy of the others. The candidate head node that has the highest residual energy sends *FinalChainHead_MSG* to inform its neighboring candidate cluster heads. If two candidate head nodes have the same residual energy, then the candidate node with the smaller *ID* sends *FinalChainHead_MSG*.

If candidate head node s_j is located in the competition range of candidate head node s_i and receives *FinalChainHead_MSG* from s_i, s_j will give up and send *Resignation_MSG* which contains the *ID*s to all candidate head nodes in its competition range, as shown in Figure 5.

If candidate head node s_i receives *Resignation_MSG* from candidate head node s_j and s_j belongs to competition range of s_i, s_i will remove s_j from its cluster group, as shown in Figure 6.

After the cluster head chain has been selected, sleeping nodes wake up and each final cluster head broadcasts a *ChainHead_ADV_MSG* containing *HeadNode_id*. Eventually, each normal node selects the cluster to which it will belong for this round (this decision is based on the received signal strength).

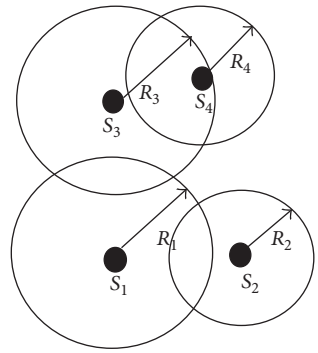


Figure 4: Competition radius among candidate cluster heads.

The larger the received signal strength is the closer the cluster head will be selected, thus, with minimum transmission power required for normal nodes to communicate with cluster head. This is carried out by normal nodes sending *ChainMember_MSG* which contains *ID*.

4.2. Chain Construction Phase. In each cluster group, chain construction starts with the cluster head node creating a

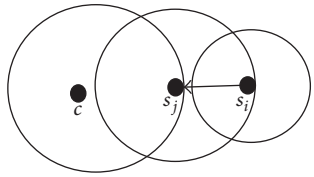


Figure 5: Node s_j receives *FinalChainHead_MSG*.

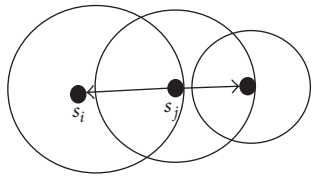


Figure 6: Node s_i receives *Resignation_MSG*.

routing table and sending it to the closest node, which then joins the chain first. The routing table contains a list of all of the nodes belonging to the cluster and the target node S_{FS} (the node at the farthest distance from the cluster head). The list of nodes is ranked based on how near they are to the respective cluster head.

During chain construction, each intermediate node forwards the request only to the neighbors that are farther away from the cluster head node than itself and have the minimum distance between each other; thus, at node s_i, the request is only sent to neighbor s_j, which satisfies $d(s_i, \mathrm{CH}_s) \leq \min(d(s_j, \mathrm{CH}_s))$, where $d(s_i, \mathrm{CH}_s)$ denotes the distance between node s_i and its respective cluster head node (CH_s), and s_i and s_j are the intermediate nodes between S_{FS} and CH_s, respectively.

In lines 1–15 of Algorithm 2, when a node receives a request packet, it checks whether it is the target node or not. If it is not the target (the farthest node from respective cluster head), it sends the request to a node that has not yet joined the chain and has the smallest distance to the cluster head.

During the chain construction process, if two nodes in the same cluster group cannot reach each other, a node can change its transmission power. The chain construction process terminates when the target node has been discovered.

Figure 7 shows the chain construction steps and data transmission between the nodes along the chain. After the chain construction process terminates and suboptimal multihop routes over the cluster heads have been discovered, the cluster head gathers sensing data within the cluster where each node receives data from a neighbor. The cluster head also fuses them with its own data, forwards them to another neighbor along the chain, and so on until the cluster head receives the fused data.

EEUCLC performs data fusion at each node except node S_{FS} located at the end of chain. Figure 8 shows a flow chart of cluster formation.

4.3. Intercluster Multihop Routing. In order to minimize the overhead, we have to ensure that the time duration of the data transmission phase is longer than the time taken by the clustering protocol to cluster the network. We utilize the advantage of the long data transmission phase to build an energy-aware intercluster multihop routing protocol by considering the residual energy of and distance to the neighboring cluster head nodes.

Since the transmission phase will be long enough to send as much data as possible to the base station, we want to ensure that there is more than one route between the cluster head nodes and the base station used for communication. Frequently using the same route between cluster head nodes leads to depleting their energy levels along the path. It leads to vast variation in the energy levels of the nodes and finally causes network partitioning. Hence, we have designed an energy-aware multihop routing protocol for the intercluster communication.

Using EEUCLC, we try to find suboptimal routes between cluster head nodes to the base station and use these routes randomly for each data transmission. In lines 1–13 of Algorithm 3, after cluster head nodes have been identified for the current round, each s_i broadcasts a message across the networks at a specific power level containing residual energy, distance to the base station, and *HeadNode_id*.

We also present a threshold *Tx_max* into the proposed intercluster multihop routing protocol so that if the cluster head distance to the base station is smaller than or equal to *Tx_max*, it directly transmits its data to the base station. When the cluster head node s_j receives broadcasted packets, it checks the distance. If a node s_j has a smaller distance to the base station than node s_i, node s_j discards the received packet. Otherwise, the node s_j selects suboptimal cluster heads to forward its data from the cluster head node candidates.

The suboptimal cluster head nodes list is defined as $s_j.\mathrm{SCH} = \{\mathrm{HE}_{\mathrm{residual}}(\min(d(s_j, s_i)))\}$, where SCH is defined as a set of suboptimal neighboring cluster head nodes which have higher residual energy and require the minimum transmission power for communication, $\mathrm{HE}_{\mathrm{residual}}$ denotes the higher residual energy of group cluster head nodes, and $\min(d(s_j, s_i))$ denote the minimum distance between the group of head nodes.

After each cluster head has chosen relay nodes, the former randomly selects nodes for each data transmission occasion from the candidate cluster headset to relay data to the base station. Figure 9 shows the suboptimal routes construction steps and data transmission between cluster heads.

5. Protocol Analysis

This section shows the analysis of the protocol described in Algorithms 1 and 2. Since the cluster head selection process is by message passing, we first discuss its message complexity.

Lemma 1. *The message complexity of the cluster formation algorithm is* $O(N)$.

Proof. At the beginning of the cluster formation, $N \times G$ candidate cluster heads are selected and each one of them broadcasts *CompeteChainHead_MSG*. Then each of them makes decision to be cluster head or not by broadcasting *FinalChainHead_MSG* or *Resignation_MSG*. Suppose k cluster

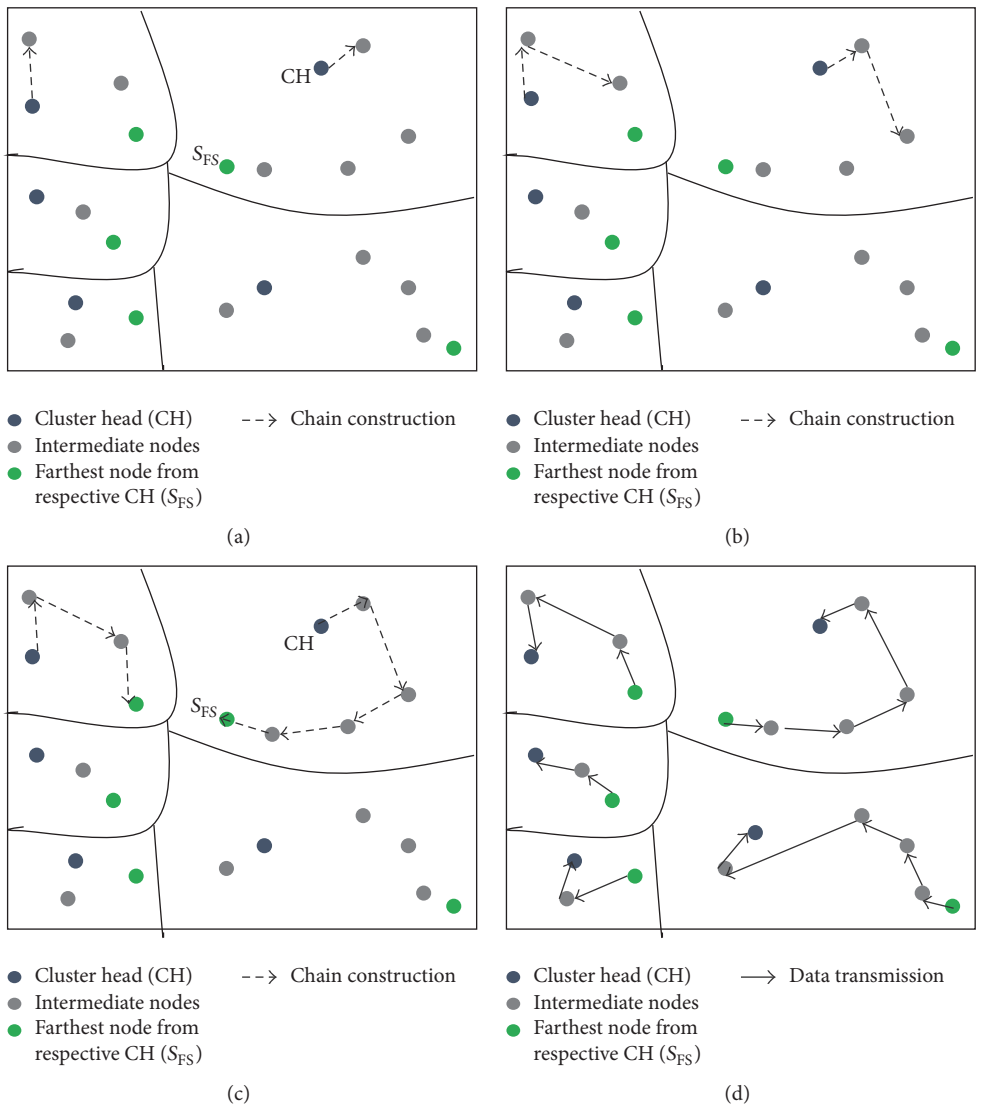


Figure 7: Chain construction and data transmission between nodes: (a) the closest node to the cluster head joining the chain first, (b) the intermediate nodes joining the chain, (c) termination of chain construction, and (d) relaying data to the designated cluster head.

heads are selected, they send out k *ChainHead_ADV_MSG*, and then $(N - k)$ ordinary nodes transmit $(N - k)$ *ChainMember_MSG*. After that $N - 1$ nodes send *Chain_construction_MSG*. Thus the messages are up to $2N \times G + k + N - k + N - 1 = 2N(G + 1) - 1$ for cluster formation stage per round. Lemma 1 shows the message overhead of EEUCLC is small. □

Lemma 2. *The time complexity for the cluster head selection algorithm is* $O(N)$.

Proof. EEUCLC protocol takes a processing time of at most $2N \times G$ among candidate cluster heads which declared their willingness to be cluster heads. This happens if more than one candidate cluster head is in the range of competition radius of each other. For the opposite, its takes at most $N \times G$ time. In both cases for sleeping nodes to join cluster it takes at most $k - N$ processing time. Therefore, the total processing complexity is $O(N)$. □

6. Simulation Results

We evaluated the performance of the proposed EEUCLC protocol using Matlab Simulation tools [27] to perform several extensive computing experiments. We also compared our simulation results with LEACH, improved LEACH, and EEUC, in terms of energy consumption by the cluster heads, network lifetime, and the number of nodes alive.

For our experiments, we simulated an environment where sensor nodes were uniformly deployed in the sensing field and the base station was located outside the WSNs. All sensor nodes periodically sensed events and forwarded data packets to the base station. The energy model adopted in our simulation is the same as that used in Heinzelman et al. [26].

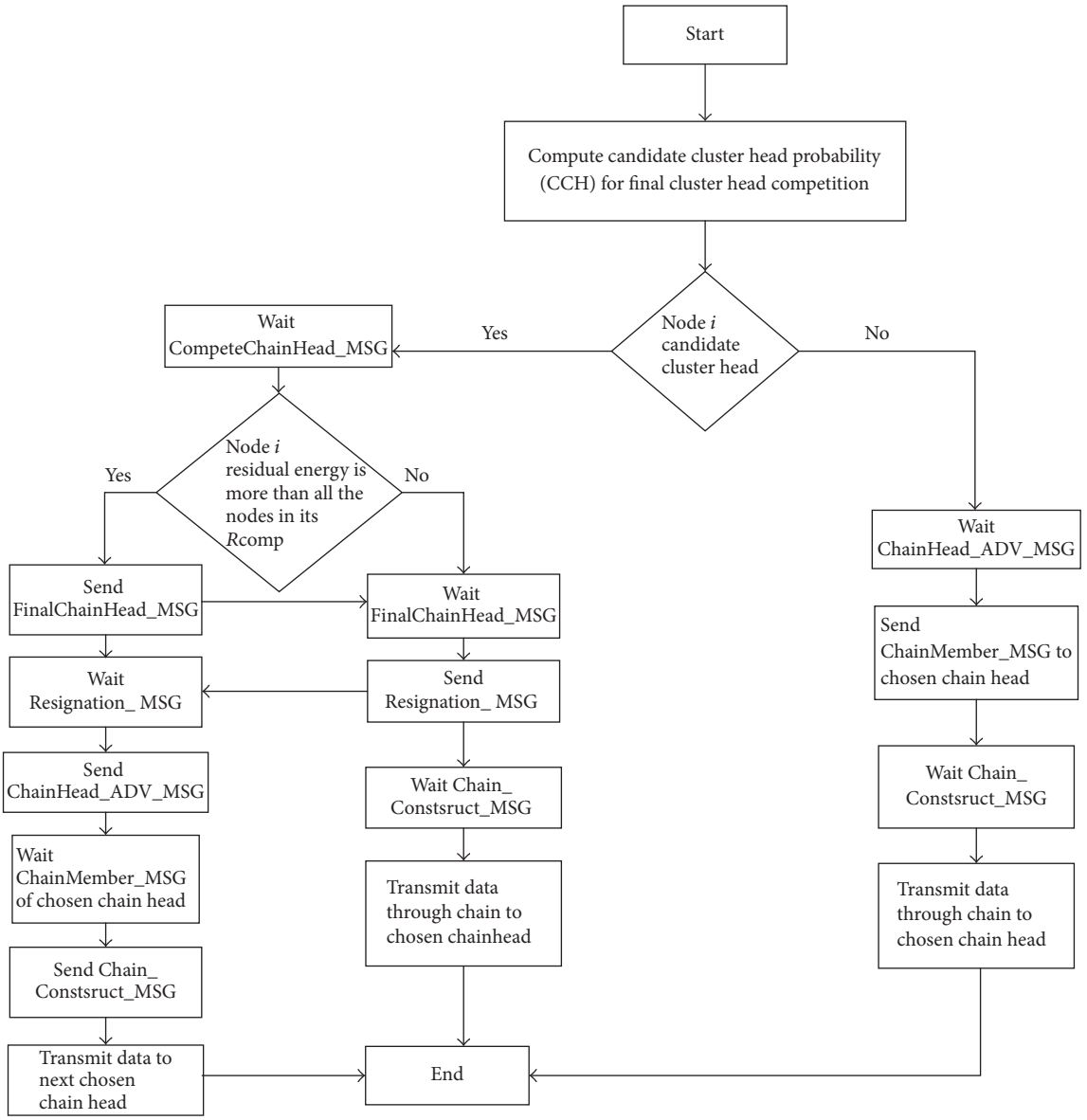


Figure 8: An overview of the EEUCLC cluster formation.

The values of the simulation parameters are given in Table 1. We set a to 0.5 in (3), R^0_{comp} to 90 m, c to 0.5 in (4), and Tx_max to 130 m. For each simulation scenario, the results are reported as the average value of 20 runs.

6.1. Energy Efficiency. In this section, we examine the energy efficiency of EEUCLC. Figure 10 presents the amount of energy spent by cluster heads derived from the EEUCLC, LEACH, improved LEACH, and EEUC schemes. The x-axis indicates 20 sampled rounds of simulation while the y-axis indicates the amount of energy spent by the cluster heads.

The results show that the energy consumed by cluster heads per round using EEUCLC was much smaller than other schemes. This is because cluster heads of LEACH and improved LEACH communicated their data directly to the base station, which consumed more energy, and because the cluster heads using LEACH were not uniformly distributed throughout the network due to the random selection of cluster heads.

Furthermore, in EEUC, the energy consumption of the cluster heads is higher than EEUCLC because EEUC tried to find the best single route between cluster head nodes and the base station and sent all aggregated data along this path, which drained the energy of the cluster heads along this path. In EEUCLC, the cluster heads transmitted their data to the base station via multihop and tried to find suboptimal routes between the cluster head nodes and the base station and transmitted data through the randomly selected suboptimal routes; thus a considerable amount of energy was saved.

We also investigated the energy efficiency of the proposed protocol by examining the network lifetime. Figure 11 shows

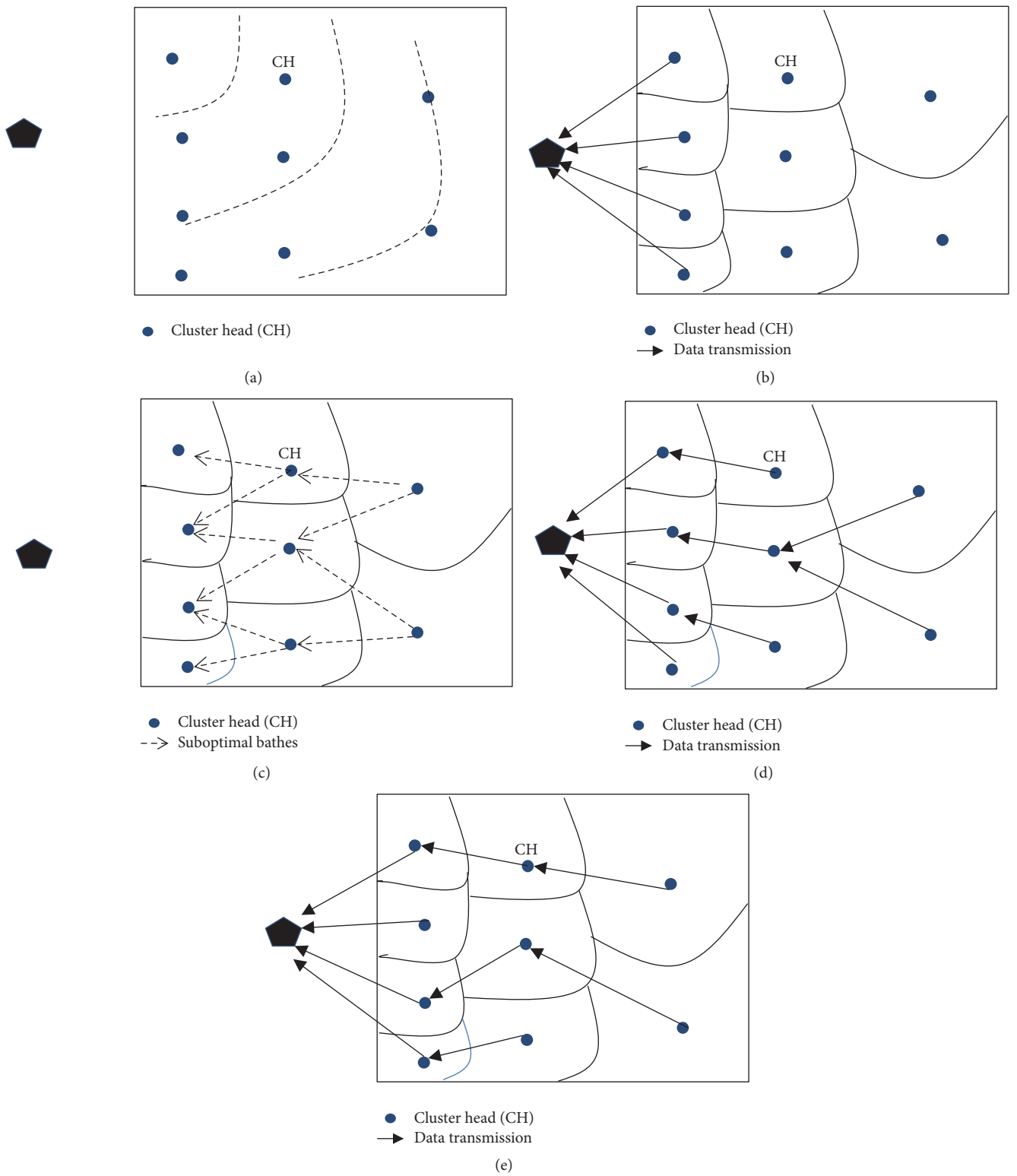


Figure 9: Suboptimal multihop routing: (a) message propagation, (b) the distance from the cluster head (CH) to the base station smaller than the threshold distance, (c) suboptimal routes selection, and (d) and (e) random selection for the relaying route.

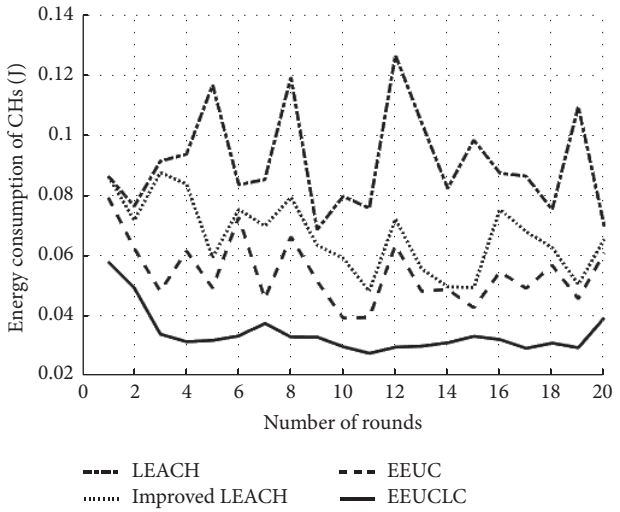


Figure 10: The amount of energy spent by cluster heads.

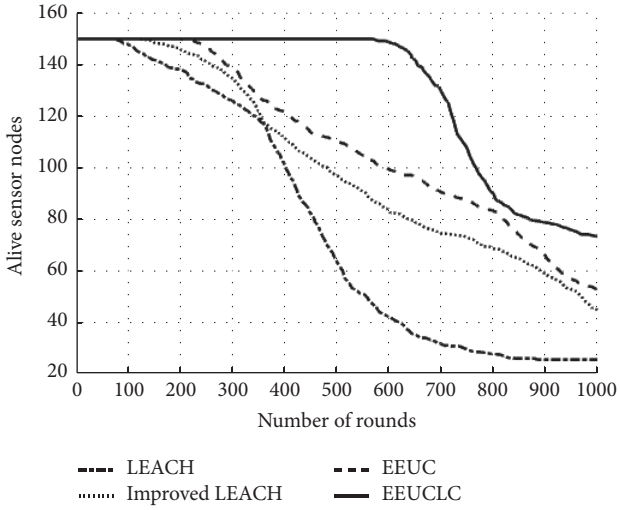


Figure 11: The time until the first node dies.

(1) On receiving *ChainHead_ADV_MSG* from cluster head
(2) *ChainMember_Msg*(*ID*)
(3) **if** cluster_head = True **then**
(4) *previous_node* ← *CH*
(5) **for** each node s_i in cluster **do**
(6) **if** $s_i \neq S_{FS}$ **then**
(7) **if** $d(s_i, \mathrm{CH}_s) \leq \min(d(s_j, \mathrm{CH}_s))$ **then**
(8) *previous_node.SendMessage*(s_i)
(9) *previous_node* = s_i
(10) **end if**
(11) **else**
(12) **EXIT**;
(13) **end if**
(14) **end for**
(15) **end if**

Algorithm 2: Chain construction phase.

(1) s_i.*Broadcast_MSG*(E_{residual}, *d*.BS)
(2) *Chain_Member_Msg*(*ID*)
(3) **if** *d*.BS < *Tx_max* **then**
(4) *message.send*(BS)
(5) **end if**
(6) **for** each s_j **do**
(7) On receiving *Broadcast_MSG*(E_{residual}, *d*.BS) from s_i
(8) **if** $d(s_j, \mathrm{BS}) > d(s_i, \mathrm{BS})$ **then**
(9) $A_r \leftarrow s_j.select(\min(d(s_i, s_j)))$
(10) $A_r \leftarrow select(\max(E_{\text{residual}}, A_r))$
(11) $s_j.SendNext(A_r.random())$
(12) **end if**
(13) **end for**

Algorithm 3: Intercluster multihop routing.

Table 1: Simulation parameter.

Parameter	Value
`Network coverage`	(0, 0)–(200, 200) m
`Basestation location`	(100, 250) m
`Initial energy`	1 J
`N`	150
E_{elec}	50 nJ/bit
ϵ_{fs}	10 pJ/bit/m^2
ϵ_{mp}	0.0013 pJ/bit/m^4
d_o	87 m
`Data packet size`	2000 bits
E_{DA}	5 nJ/bit/signal

the time until the first node died, while Figure 12 shows the time until the last node died.

EEUCLC clearly improved not only the time until the first node died, but also the time until the last node died compared to other schemes. With a sensing field of size 200 m × 200 m, the graphs show that sensor nodes in EEUC, improved LEACH, and LEACH nodes died earlier compared to EEUCLC because their sensor nodes became thinly scattered and thus consumed more energy to convey their data over the long distance to their respective cluster.

Moreover, some cluster head nodes in improved LEACH and LEACH consumed more energy to convey their data directly over a long distance to the base station. Since cluster heads consume energy based on the number of nodes within a cluster, the cluster heads in EEUC with low energy located far away base station died earlier. This is because EEUC is an unequal clustering method, where clusters far away from base station have a larger size than clusters near the base station. EEUCLC overcame these issues by convening data to the respective cluster head using chain schema, and so nodes having low residual energy at a long distance from the base station had a low probability when competing for the final cluster head.

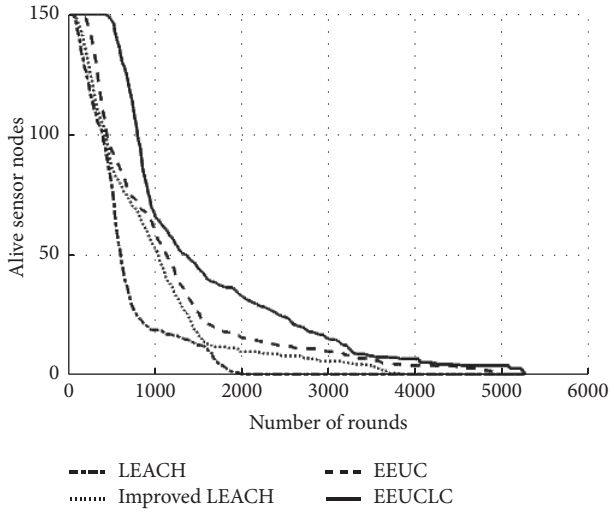


Figure 12: The time until the last node dies.

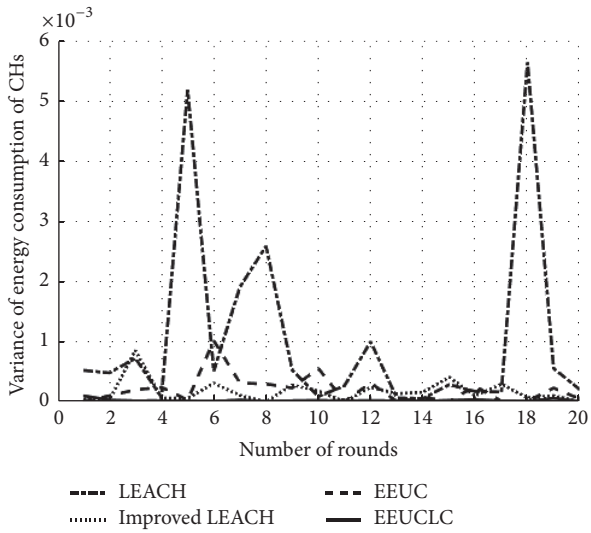


Figure 13: The variance of amount of energy spent by cluster heads.

Figure 13 displays the variance of energy consumption by cluster heads in 20 sampled rounds. It shows that EEUCLC balanced the energy consumption between cluster heads better than other schemes. This is because the cluster heads distribution in LEACH and improved LEACH was not controllable; thus the number of nodes within clusters varied too much, while, in EEUC, some cluster heads relayed all aggregated data received by multiple cluster heads. Hence, some cluster heads consumed more energy than others. EEUCLC avoided the hot spot problem by utilizing its suboptimal multihop routing mechanism.

Figure 14 exhibits the number of rounds until 1%, 25%, 50%, and 100% of nodes died in a 200 m × 200 m field. In this field size, sensor nodes started to die at a more rate after about 20% of the nodes had died. This was due to the distances between the nodes which became larger, and nodes had to become leaders more often, causing the energy to deplete quickly. EEUCLC achieved approximately 1x the number of

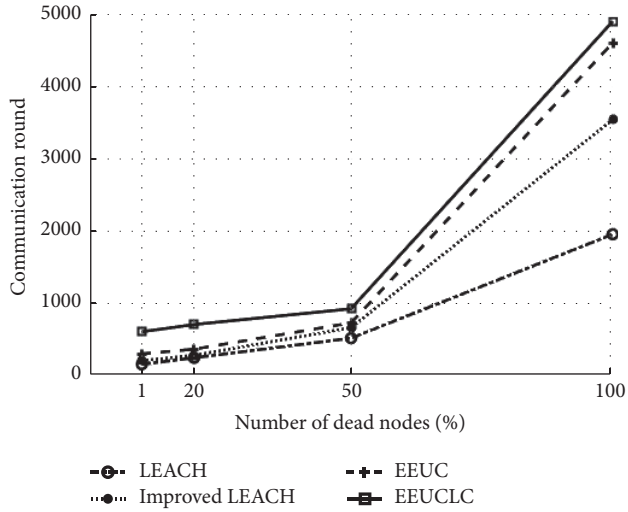


Figure 14: Performance results for a 200 m × 200 m network with initial energy 1 J/node.

rounds compared to EEUC when 1%, 20%, 50%, and 100% of nodes died for a 200 m × 200 m network, while compared to improved LEACH achieving approximately 2x the number of rounds. Last, compared to LEACH, EEUCLC achieved approximately 3x the number of rounds.

7. Conclusion

In this paper, we proposed EEUCLS, a novel multiple chain construction protocol capable of data aggregation and conveying data to a base station for WSNs. The proposed protocol provides balancing the energy consumption among the cluster heads, preserving energy consumption of the nodes within the cluster, and reducing accelerated energy consumption of cluster heads. To support multihop routing in EEUCLS, we presented an unequal chain length clustering and a suboptimal multihop routing protocol to alleviate the burden on the cluster heads and prolong the lifetime of the sensor network.

In addition EEUCLS can also work in a very variable sensor networks domain, for example, integrated navigation system with various sensors (GNSS, GPS, and Speed Log) because it can secure the availability, consistency, and synchronization of the information from the various sensors. Simulation results show that our unequal clustering protocol clearly improved the network lifetime and balanced the energy consumption by the cluster heads compared to LEACH, EEUC, and improved LEACH.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This material is based upon work supported by the Ministry of Trade, Industry & Energy (MOTIE, Korea) under Industrial Technology Innovation Program no. 10063496,

"Integrated Navigation System Development Based on Virtual Platform."

References

[1] T. Chen, "Smart grids, smart cities need better networks," *IEEE Network*, vol. 24, no. 2, pp. 2-3, 2010.

[2] A. Manjeshwar and D. P. Agrawal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," in *Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS '01)*, vol. 1, pp. 2009–2015, IEEE, San Francisco, Calif, USA, 2001.

[3] J. Huang, Q. Duan, C.-C. Xing, and H. Wang, "Topology control for building a large-scale and energy-efficient internet of things," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 67–73, 2017.

[4] J. Huang, Y. Meng, X. Gong, Y. Liu, and Q. Duan, "A novel deployment scheme for green internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 196–205, 2014.

[5] Y. Lim, H.-M. Kim, and S. Kang, "A design of wireless sensor networks for a power quality monitoring system," *Sensors*, vol. 10, no. 11, pp. 9712–9725, 2010.

[6] J. Ahmad, A. S. Malik, M. F. Abdullah, N. Kamel, and L. Xia, "A novel method for vegetation encroachment monitoring of transmission lines using a single 2D camera," *Pattern Analysis and Applications*, vol. 18, no. 2, pp. 419–440, 2015.

[7] D. Curiac and C. Volosencu, "Urban traffic control system architecture based on wireless sensoractuator networks," in *Proceedings of the 2nd International Conference on Manufacturing Engineering, Quality and Production Systems (MEQAPS '10)*, Constantza, Romania, 2010.

[8] M. Ceriotti, M. Corrà, L. D'Orazio et al., "Is there light at the ends of the tunnel? Wireless sensor networks for adaptive lighting in road tunnels," in *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN '11)*, pp. 187–198, Chicago, Ill, USA, April 2011.

[9] Y.-F. Chung and C.-H. Liu, "Design of a wireless sensor network platform for tele-homecare," *Sensors (Switzerland)*, vol. 13, no. 12, pp. 7156–17175, 2013.

[10] J. Ko, T. Gao, and A. Terzis, "Empirical study of a medical sensor application in an urban emergency department," in *Proceedings of the 4th International ICST Conference on Body Area Networks (BODYNETS '09)*, Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, April 2009.

[11] X. Mao, X. Miao, Y. He, X.-Y. Li, and Y. Liu, "CitySee: urban CO_2 monitoring with sensors," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 1611–1619, Orlando, Fla, USA, March 2012.

[12] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[13] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *IEEE Wireless Communications*, vol. 11, no. 6, pp. 6–28, 2004.

[14] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Siences (HICSS '00)*, p. 10, January 2000.

[15] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings of the IEEE Aerospace Conference*, vol. 3, pp. 1125–1130, Big Sky, Mont, USA, March 2002.

[16] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.

[17] C. F. Li, M. Ye, and G. Chen, "An energy-efficient unequal clustering mechanism for wireless sensor networks," in *Proceedings of the 2nd IEEE International Conferenceon Mobile Ad hoc and Sensor Systems (MASS '05)*, Washing ton, DC, USA, 2005.

[18] M. Demirbas, A. Arora, V. Mittal, and V. Kulathumani, "A fault-local self-stabilizing clustering service for wireless ad hoc networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 9, pp. 912–922, 2006.

[19] R. C. Shah and J. M. Rabaey, "Energy aware routing for low energy ad hoc sensor networks," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '02)*, vol. 1, pp. 350–355, IEEE, March 2002.

[20] E. Ever, R. Luchmun, L. Mostarda, A. Navarra, and P. Shah, "UHEED - An Unequal Clustering Algorithm for Wireless Sensor Networks," in *Sensornets*, Rome, Italy, Feburary 2012.

[21] P. Ren, J. Qian, L. Li, Z. Zhao, and X. Li, "Unequal clustering scheme based LEACH for wireless sensor networks," in *Proceedings of the 4th International Conference on Genetic and Evolutionary Computing, ICGEC 2010*, pp. 90–93, chn, December 2010.

[22] J. Yu, Y. Qi, G. Wang, Q. Guo, and X. Gu, "An energy-aware distributed unequal clustering protocol for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 7, no. 1, Article ID 202145, 2011.

[23] S. Lee, H. Choe, B. Park, Y. Song, and C.-K. Kim, "LUCA: an energy-efficient unequal clustering algorithm using location information for wireless sensor networks," *Wireless Personal Communications*, vol. 56, no. 4, pp. 715–731, 2011.

[24] J. Yang and D. Zhang, "An energy-balancing unequal clustering protocol for wireless sensor networks," *Information Technology Journal*, vol. 8, no. 1, pp. 57–63, 2009.

[25] W. B. Heinzelman, *Application specific protocol architectures for wireless networks [Ph.D. dissertation]*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2000, https://pdfs.semanticscholar.org/3b86/506f0562c1453461509c5dc43359fe-908db5.pdf.

[26] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, 2002.

[27] The matworks matlab and simulink for technical computer: http://www.mathworks.com/.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 1058081, 9 pages
https://doi.org/10.1155/2017/1058081

Research Article

A Cyberphysical System Based Mass-Customization Approach with Integration of Industry 4.0 and Smart City

Mehmet Karaköse and Hasan Yetiş

Department of Computer Engineering, Fırat University, Elazığ, Turkey

Correspondence should be addressed to Hasan Yetiş; h.yetis@firat.edu.tr

Received 31 March 2017; Revised 24 July 2017; Accepted 6 August 2017; Published 30 August 2017

Academic Editor: Danda B. Rawat

Smart city is a city which is designed to meet the people's demands. In addition to use of sources efficiently, trends of people are also a need that smart city should meet. Buying personalized products in a cheap and fast way is a demand of people of today. Mass customization, which is defined as the personalization of products, achieves making the tailor-made products cheaper. In this study, we propose a new approach for mass customization with the integration of smart retail and smart production. With removing the operators and actualizing the progress autonomously, it is aimed to reduce the waiting time of customers. Because less waiting time means that there are more mass-customization customers, and this is expected to increase the popularity of mass customization. Thus, reducing wastes and increasing productivity are aimed. This study also constitutes the infrastructure that enables a production system to autonomously perform all stages from order to delivery. With the given scenarios, challenges and advantages of desired approach are discussed.

1. Introduction

Smart city is relatively new and developing concept among the other city concepts such as talented city, wired city, and ecocity. Because it contains a part of other city concepts, smart city is a wide concept [1]. That is why it is possible to encounter various smart city projects in different types and fields, and it is difficult to make a specific smart city definition [2]. In general, the smart city is a city equipped with methods which are developed to meet increasing population needs more effectively and efficiently [3]. The world has become a very crowded place with an increase in population, so crowded countries such as China have developed some measures for population limitation. The management of city resources that remain constant or limited over the growing population has become very important. These resources not only are natural sources such as air, energy, and terrain but also are human-generated sources such as roads, parking areas, and buildings. Furthermore, the needs do not always arise from the necessity; the things that people desire are also a need. So, smart city should meet the demands of people. Minimizing human factor and making work easy are among the main subjects and needs of humanity since the first era. Technology has evolved in this way.

When we look at the development of industry, the first three stages of industrial revolution have actualized by using steam-powered machines, electricity-powered machines, and programmable microprocessors in manufacturing, respectively. Today, 4th industrial revolution which is named Industry 4.0 is actualizing using cyberphysical systems (CPS) in Industry [4]. Industry is a part of city and there are common interests between Industry 4.0 and smart city concepts. They both aim to minimize human interaction and the energy consumption. Industry 4.0, which aims at totally man-free production, is more energy efficient and more manageable [5]. Common interests of smart city concepts and Industry 4.0 benefits are shown in Figure 1.

In the preindustrial era, the production type was tailor-made productions done by skilled craftsmen, which is called craft production. While the industrial revolution and production lines bring together one type of mass production, such special productions continue to exist. However, craft production is not a widely preferred method because the prices are visibly different. A faster and cheaper production

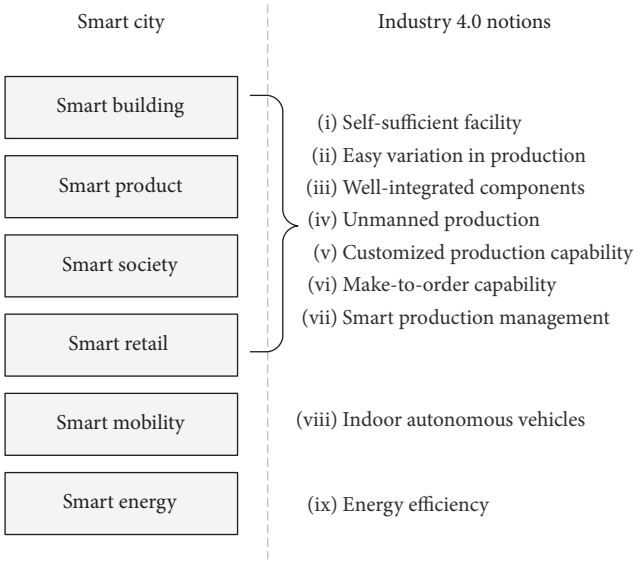


Figure 1: Smart city components and their relations with Industry 4.0 notions [6].

model that can produce limited-count different variations of a product, which is called lean production, is introduced after mass production [5]. According to this method transport, inventory, motion, waiting, overprocessing, overproduction, and defects (TIMWOOD) are main sources of wastes in manufacturing and they should be handled for more effective manufacturing. Later, producing customer special products at a cheaper price in the factory environment to meet the customer's needs reveals the concept of mass customization. In summary, after the industrial revolution, mass production, which can produce one type of production, lean production, which enables producing predefined product types with some adjustments on the system within minutes, and mass customization, which is actualized upon the demand of the customers, were brought out. Although it seems impossible to customize production with production bands, it is possible to achieve this with 4 factors. These key factors which make mass customization easy are as follows:

(i) Flexibility
(ii) Connectivity
(iii) Modularity
(iv) Organization

A CPS infrastructure ensures the key factors which are required for mass customization such as flexibility, modularity, and Internet. Furthermore, because it is possible to fix the system with an intervention to the cloud and to check the data in the system, CPS is superior in terms of integration with other systems. So, in order to achieve mass customization, using CPS and Industry 4.0 is a suitable way. Cyberphysical systems consist of three main layers. These are cyber, physical, and communication layers. As the name implies, the physical layer contains sensors that measure physical environment values and actuators that are in contact with the physical environment. In the cyber

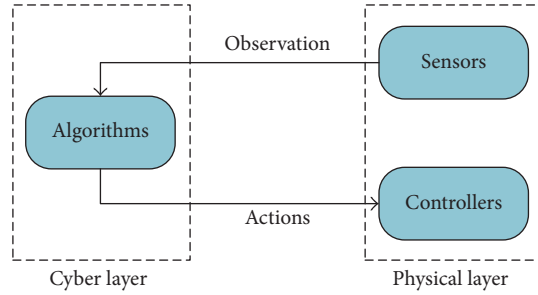


Figure 2: Cyberphysical system structure [7].

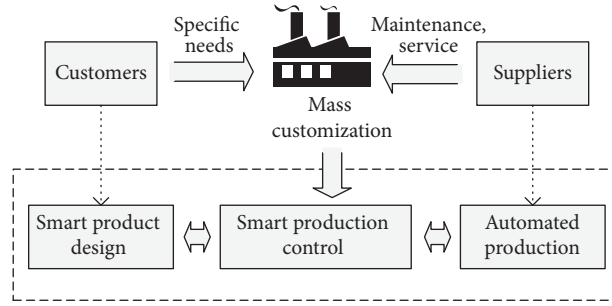


Figure 3: The block diagram about actualizing of mass customization.

layer, the data coming from the physical environment are processed and necessary algorithms are created to produce appropriate behavior. Computational operations take place in the cyber layer. The communication layer is responsible for transferring the data from the physical layer to the cyber layer and transferring the commands from the cyber layer to the physical layer. Basic structure of CPS is given in Figure 2.

Mass customization, which forms a part of this study, can be actualized in 3 ways. The first one is the make-to-stock method in which a product is manufactured uniformly; then the user can customize the product according to his/her needs. The second one is the assemble-to-order method that is achieved by combining the product's previously produced modules in line in accordance with the customer's request. The last one is the make-to-order method in which the production starts upon customer order. Mass customization allows the production of special products at low cost in terms of color, function, feature, or style. Since the tailor-made production requires to carry out the production on order, it removes the problems such as stockpiling, opening a store, or surplus stocks. Because of profitable in terms of customers and producers, it is one of the big trends of today's manufacturing. For these reasons, it is being implemented successfully by the pioneer companies such as Nike, Dell, and Mymuesli. It is also seen possible making special medicine at production line for each patient in near future. A block diagram representing actualizing of mass customization is given in Figure 3.

Although mass customization has advantages, the supply time is also increased compared to the normal products and this partly reduces its attractiveness. For being useful,

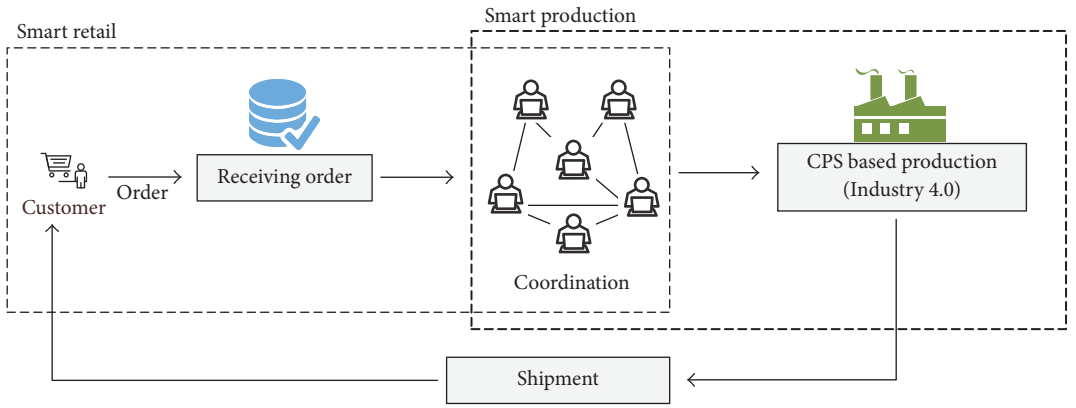


Figure 4: Integration of smart retail and Industry 4.0 for mass-customized production.

essentially a customized product should not cost twice or triple the price of a mass-produced item, and it should not take twice as long to receive the customized product either. People of today demand minimizing human interaction, buying tailor-made productions in a cheap way, and actualizing things faster. In this paper a new mass-customization approach is proposed in order to meet the demands of era. The main aim of this study is making the time shorter from order to delivery by minimizing human interaction for mass-customized products. Most of the procurement period goes through the processing of the received order and the adjustment of the configurations. In this study, a method has been proposed to shorten the CPS-based supply period, including smart trade, which is part of Industry 4.0 and Smart City, for mass customization. Cyberphysical systems are an infrastructure that removes the necessity of microcontrollers to be at physical components with the development of Internet technologies. As they are more advantageous than classical embedded systems in terms of flexibility, reliability, and controllability, classical embedded systems have begun to be changed with cyberphysical systems [9].

2. Materials and Methods

Buying tailor-made products in a cheap and fast way is a demand of today's people. Cheaper tailor-made products can be produced thanks to the mass customization. This paper aims at reducing the duration of supply of mass customization. Most of the time spent for production is caused by human factor. We proposed an approach that removes the operators between order and production events in order to make the process faster. Internet-aided order-to-production steps and operators that coordinate order and production steps are shown in Figure 4.

Today, you can perform bank transactions, pay bills, and buy something thanks to the Internet and technology. The Internet removes the need of human factor. Therefore, the scenario is built on online shopping. According to the scenario, user is able to customize his/her production via the Internet as such in current applications. After specialization of his/her product, autocalculated price is offered to customer. If he/she agrees, then payment and order processes are started. When the order delivers to system, it is verified automatically by the application at server, and then it is posted to subsystem which coordinates the totally man-free production aimed factory. So, the operators that coordinate these two systems are removed.

Removing the coordination unit from the system has some challenges. For example, when the coordination is actualized by humans, they can verify that their message is sent successfully. But machines need some more effort to do it. Some precautions should be taken to prevent data loss. Hence, the two subsystems should handshake to confirm that the order is delivered to other system. Basically, the progress of order is as follows. The system takes the order and adds it to process queue. Taking order and adding it to process queue is actualized in one transaction in order to prevent data loss. So, when the order is added to process queue, order status on server, which user can access, is updated. An approximate production date according to queue status is also sent to the server. Any change on production status triggers the server to update; hereby customer can track his/her production status real-time. Furthermore, with use of the process queue, it is possible to give priority to customers who pay more for a faster supply. This method can bring new ways of earning for companies. The basic block diagram of the approach is given in Figure 5.

Figure 5 shows a basic diagram of progress. Dotted squares represent the two different stages which are smart city and Industry 4.0 production environment. The task of order server, which is between smart city and Industry 4.0 environment, is receiving the customized orders through Internet and transfers them to Industry 4.0 environment safely. Diverging the stages makes the system secure, but it brings also synchrony problems together. If any exception occurred in any step, system marks the process as uncompleted, and uncompleted processes are triggered by system at certain intervals. Thus, the system works as a whole and order losses are prevented. In the factory environment, which is dotted rectangle at right in Figure 5, production process is actualized by CPS infrastructure. So, the adjustments for each special product are made by simple changes on software in cyber layer. According to CPS, the data which are obtained by sensors in physical layer are sent to computational tools in

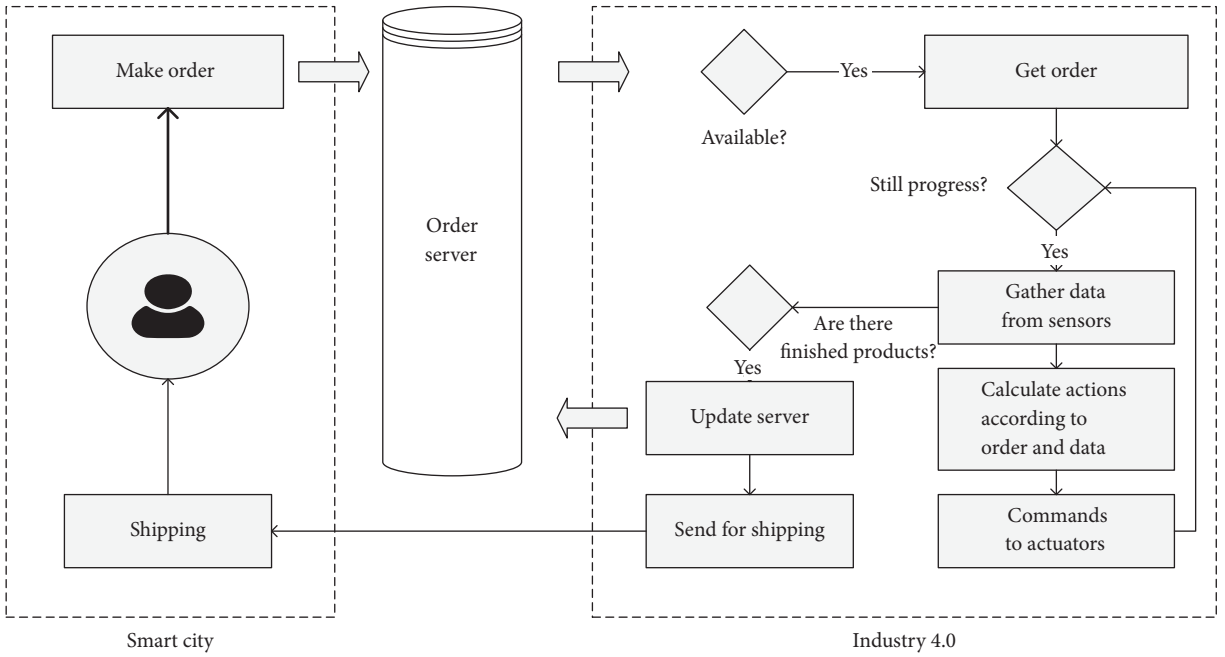


Figure 5: The basic block diagram of the proposed approach.

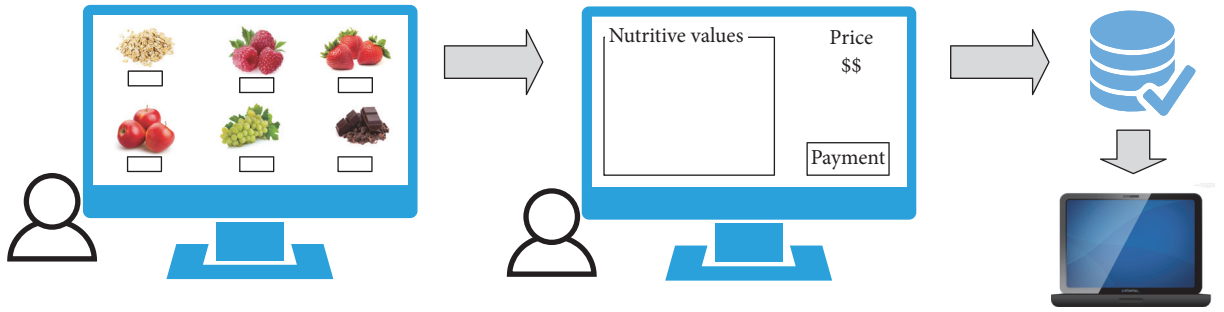


Figure 6: Customizing and making payment of crispy.

cyber layer thanks to communication technologies. The cyber layer makes the calculations according to the customer order and the sensor data to produce appropriate commands for machines in physical layer. After the production of product, it is shipped by autonomous vehicles, which is a study field of smart city [10, 11]. Hereby, the production process from ordering to shipping can be done without any human interaction. For a better understanding of the method, let us consider a firm that produces fruity crispy. For example, there can be raspberries, strawberries, apples, grapes, and chocolate in crispy. Customer is able to determine the amount of fruit and oats. Figure 6 shows the ordering and taking into the request. The request is sent to CPS main processor.

Because of their real-time communication skills, the product is ready to be produced immediately after the order. In CPS structure, the main processor configures the production platform according to orders. For example, consider that there are more than 6 orders in queue. An example situation is given by Figure 7. M0 to M7 are actuators which actualize the given movement. M0 is for rotation of the production line, M7 is for packaging, and M1–M6 are for setting the amount of fruits. The main computer has the information of all orders to be produced. So, it decides to work of actuator. For example, at the current time shown in Figure 7, $n + 3$th product needs 50-grams apple, and n +1th product does not need chocolate. So, while M6 is not activated, M4 is made active for some time. If all actuators done their work, then M0 works and the next configuration for production is loaded.

Without the operators, the production is realized. But, such a system brings main two questions, the first one is that how secure is it to connect the factory to Internet; and the second one is that in which cases running the factory is logical, because security and effective task scheduling are crucial for any large system [12].

First, establishing such a system may seem insecure, because when we connect our factory to Internet it becomes vulnerable to attacks from the Internet. An example of this occurred in October 21, 2016, as IOT devices were hacked [13]. There are studies that reduce the effect of such problem in literature [14]. But in fact, there is no real connection between user and the facility network directly in this paper. Customers can access the server which they can make orders and the

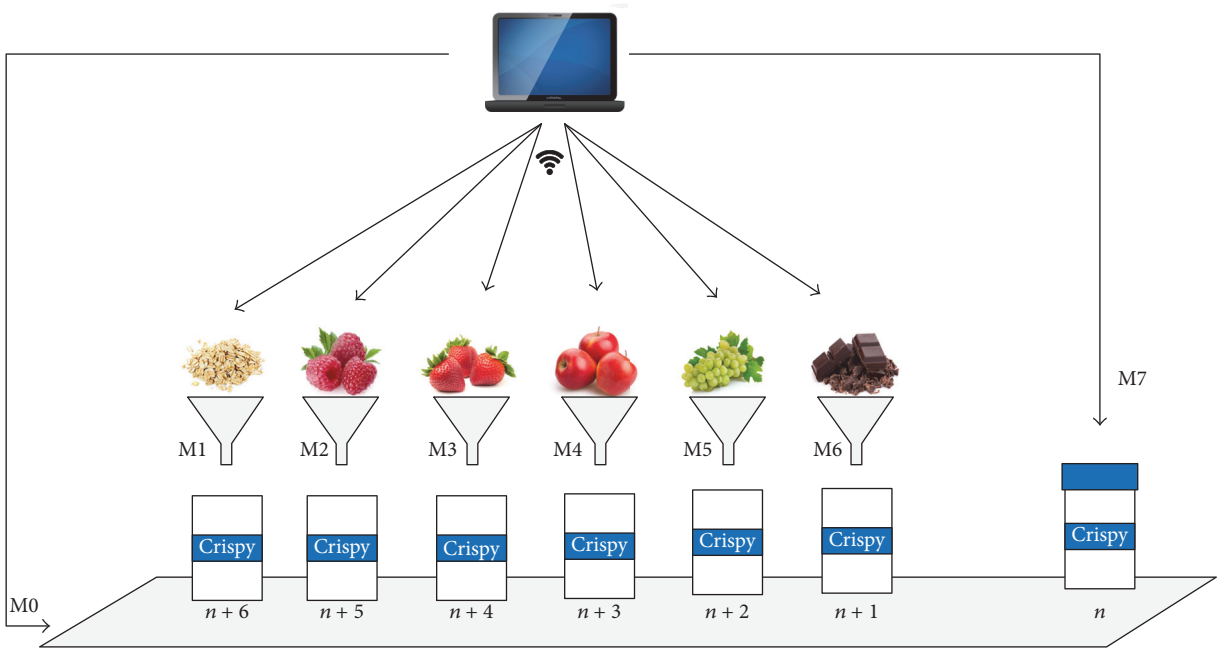


Figure 7: A CPS-based crispy production line.

server sends the order only when the order is completed. So, separating the systems which serves customers and the systems which organize the facility makes the system reliable for Internet attacks such as DDOS (distributed denial of service) [15]. The same is true for the customer.

The second problem is working the machines in factory for just a one order or a few orders. Put machines in action may require more energy and starting and stopping them often may cause much more cost. So, when machines are working, the new order is taken to process queue, and when the machines are stopped the new order is taken to waiting queue. If the order is taken while machines are in action, it means that the product will be produced soon. But if the machines are resting, it means that the order should wait for a while for achieving fulfilled condition. A block diagram for new order status is given in Figure 8.

The fulfilled condition depends on the following:

(i) Total waiting order count

(ii) Total waiting time of customers

(iii) The min amount of order for efficient working of machines

(iv) The rate of importance of customer happiness to importance of profit for companies

An algorithm to decide whether running the factory is logical or not can be created with considering the factors above. If the algorithm says it is logical then the machines, which are stopping, are starting to work. The adjustment of the production elements in the factory to the order received will be done in a completely autonomous manner. With integration of assistive technologies, system can also be used by disabled people easily [16]. Using CPS structure makes the system flexible and it also makes the system developable.

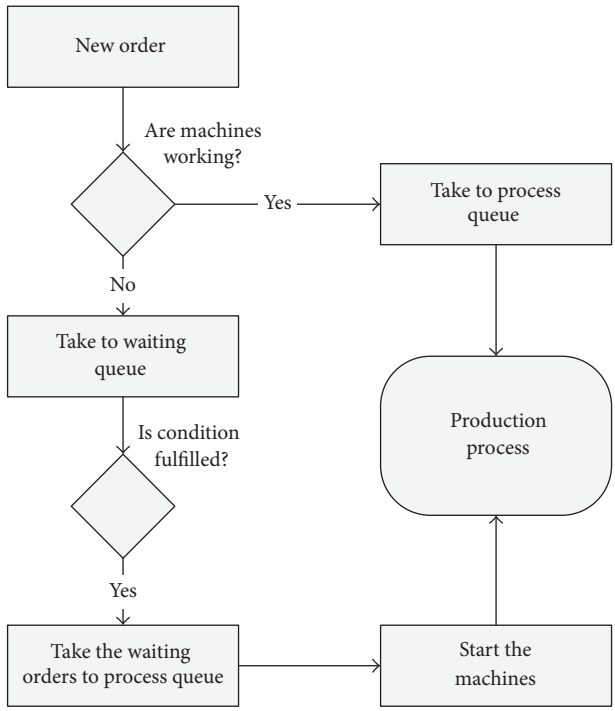


Figure 8: Add new order to queue according to machines status.

3. Results and Discussion

Most of the firms estimate their sales and make stock, and they can incur losses due to the overstock. Mass customization, whose production is started after the order, is a new way of manufacturing. Mass customization not only is helpful with firms but also provides personal products to customer.

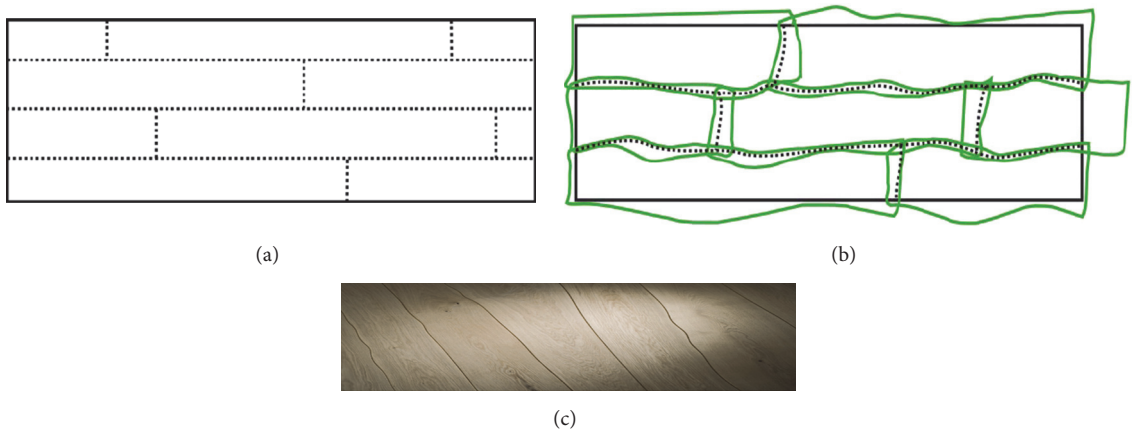


Figure 9: Rectangular floorboards and curved floorboards [8].

So, it is a production method based on win-win relationship. But the extra waiting time for buying personalized product is the biggest problem in front of the mass customization. It is the fact that machines are much faster in specific areas, they programmed. And they are prone to make less mistakes. So, human interaction is a factor which must be minimized in order to achieve faster and more accurate works. Therefore, in this study, we aimed to remove the operators between the order and production steps.

Smart city aims to achieve strong city characteristics such as economy and culture, with more efficient use of the physical infrastructure by using cybercomputations such as artificial intelligence and data analytics [18]. On the other hand, electronic platforms, which are also used for making people more active in choices about city and even country, are utilized in smart city concept [19]. Using such platforms not only make the people life easier but also provide us with data to analyze other components of smart city. So, we use Internet and mobile platforms to taking orders. With the electronically gathered information, some estimates can be done to make the production process more efficient. Furthermore, the information can be used for learning algorithms and self-modified systems can be achieved [20].

Using Internet and connecting the two different parts make the system faster. Minimizing the human factor makes the system smarter. For example, one of the firms which use smart manufacturing systems is Bolefloor [8]. Bolefloor Company achieved the production of mass customization based novel wooden design products in flexible way. They aim to save raw materials which are cropped while trying to make the floorboards rectangular, and they produce curved shaped floorboards instead of rectangular thanks to the technology they used. A sample floorboard and the raw materials for producing them are given in Figures 9(a) and 9(b), respectively. The visual image of the floorboards with curved edges is given by Figure 9(c).

While the floorboards have to be cropped straight shaped to fit each other in mass production, mass customization has removed this obligation. As seen by Figure 9, raw materials should be cropped to obtain rectangular floorboards. After the cropping process, the cropped raw materials, which is shown by the area of between green and dotted lines in Figure 9(b), turn into trash.

Producing mass-customized floors is more complex than producing rectangular floors because it is hard to assemble nonstraight edged shapes. In order to find the optimal solutions, application software and hardware should work in a harmony. Such a system depends on the similar characteristics with CPS. The block diagram of the process of mass-customized floorboard production is given by Figure 10. The figure shows us that the components of system, both 3rd part software and hardware, work in a tight interaction.

Mass customization is used in manufacturing such as vehicles, mobile devices, clothes, and footwear. It also can be used in eyeglass manufacturing. Eyeglass frames are shaped by warping the materials by machines at same ratio, so the products are limited to few models. Warping process of a frame raw material is given in Figure 11.

But with warping the raw material at different ratios, much more eyeglass models can be produced. With considering a factory which has machines that can produce eyeglass frames with different size, color, shape, and customized text and package, people are able to customize their products through the Internet and the users are supported by order customization applications to make designs suitable to produce. A sample system is given in Figure 12.

The machines in Figure 12 are set up to suit the CPS infrastructure and they can be controlled easily by the main computer in cyber layer. Main computer sends the information of what they do to machines, and machines act according to this information. The new orders which users make are added to order queue in main computer. If the factory is working or the working condition is satisfied, the orders are being produced. If the machines are stopped and the orders in queue do not maintain the working criteria, then the orders wait until the working criteria are satisfied. Working and resting time of facility can be controlled easily in cyber layer. The orders which will be produced are being processed according to algorithms, so, for example, if a person pays more for faster delivery, a priority may be given

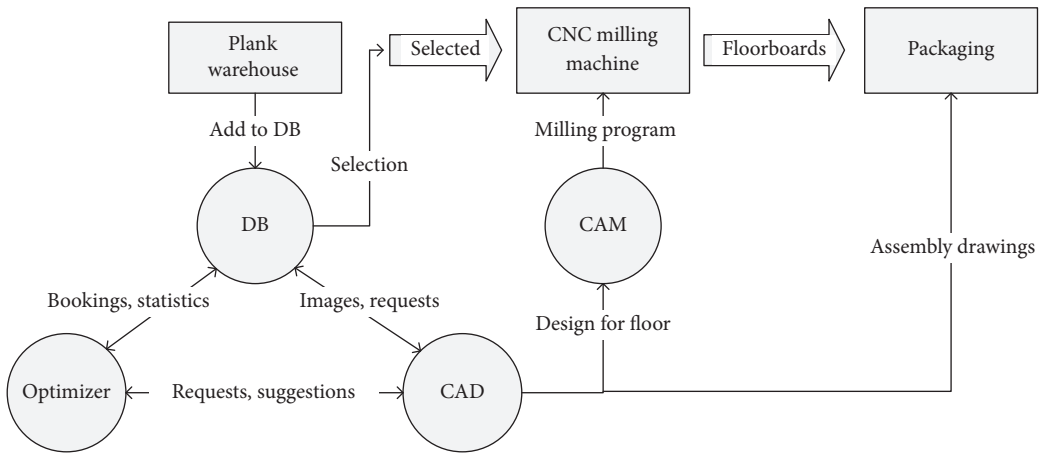


Figure 10: The production process of curved floor [8].

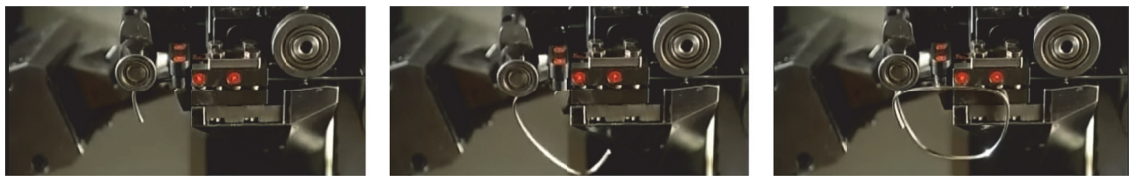


Figure 11: The warping process of eyeglass raw material.

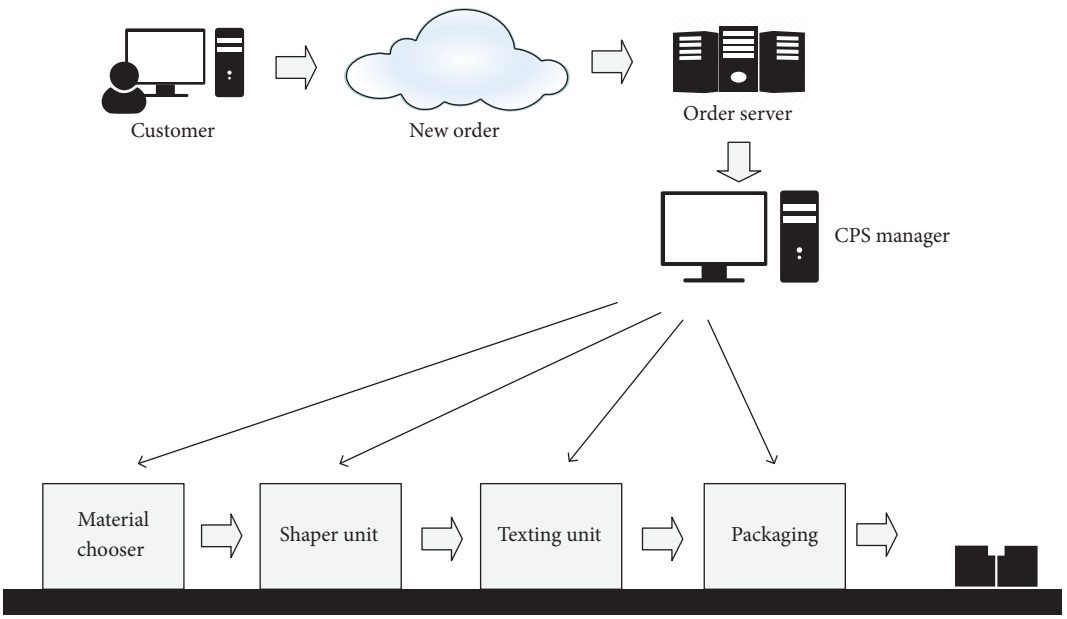


Figure 12: A sample about the eyeglass frames for proposed method.

to his/her process. And, this may provide extra revenues for firms.

Lanyman is a German mass-customized keychain company. Their prices and shipping times are given in Table 1. A person in Germany must pay 8.90€ to buy a personalized keychain. This is not problem for customers, because standard keychains are at similar prices. But the waiting time may be deterrent factor. While a person buys standard keychain at markets immediately, he/she must wait at least 5 days to get customized one. Considering the shipment time is about 2-3 days, the product needs 3–5 days to be produced. It is too much and probably more time is spent on operators [17]. If we remove them, we can produce the product much earlier. Considering there are 1000 orders and an order is produced in 30 seconds, so 30000 seconds is needed to produce all orders. It is about 8.5 hours, and it means the orders are

Table 1: lanyman.com shipping and handling conditions [17].

Country	Cost (€)	Waiting time (day)
Bulgaria	9.65	9–12
Germany	4.60	5–7
Denmark	8.60	7–9
Finland	24.70	10–13

produced in shorter than one day. Shipments are done at the end of the day, so an order needs about 1 day to be produced. When we compare it with the current method, the proposed approach is 60–80% faster. The other companies may not wait for production so much as this. So the gain changes from case to case.

Machines in a factory work according to pipeline method. So when a machine working on one product, another machine works on probably another product. This makes the system responsible for managing all machines simultaneously one by one. For example, in our sample, when raw material is selected for nth order, shaping operation is working for $n + 1$th order. At the end of the production line, order is produced and packaged and is ready to ship. Because we know each product's identical information, the location where the produced order will be shipped is also known. So the production can be shipped easily by the autonomy shipping vehicles which are a part of smart city. There is autonomy shipping system with drones for near places, within about 30 minutes, but it is seen possible for far places thanks to autonomy transport in near future.

4. Conclusions

Mass customization is one of the preferred production models today because of its structure which makes possible to produce specialized products with low cost. However, because of its order-to-product architecture, it needs more time for delivery than mass-production model. The time for ordering, manufacturing, and shipping and the time between the steps reduce mass customization's appeal. This paper presented an approach which reduces the elapsed time with integration of these steps in a secure way. For this aim, cyber city components and Industry 4.0 are combined in cyber-physical system infrastructure, so that the order, production, and delivery steps can be performed automatically and faster.

The problems occurred when trying to integrate the two different terms, such that security and efficiency are handled. An order server, which takes order from customers and transfer it to factory environment when completed, is used as middle layer between smart city and Industry 4.0 layers. It makes the system more secure against Internet attacks. As it is not efficient way to make the machines working in factory for a few orders, they should be started according to an algorithm that optimizes the parameters.

Speeding up the process of mass customization makes it more appealing and more preferable for customers. Producers will take into account the demands of customers inevitably. And the count of factories using Industry 4.0, which is more energy efficient, will increase. This contribute energy saving and will have an additive positive effect to economy and ecosystem also.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by Scientific Research Projects Coordination Unit of Firat University, Project no. MF.17.09.

References

[1] C. Manville, G. Cochrane, J. Cave et al., "Mapping Smart Cities in the EU , European Parliament's Committee," Tech. Rep., 2014.

[2] A. Cocchia, "Smart and Digital City: A Systematic Literature Review," in *Part of the series Progress in IS*, pp. 13–43, Springer International Publishing, 2014.

[3] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25–36, 2014.

[4] H. Yetis, M. Baygin, and M. Karakose, "An investigation for benefits of cyber-physical systems in higher education courses," in *Proceedings of the 15th International Conference on Information Technology Based Higher Education and Training, ITHET 2016*, Istanbul, Turkey, September 2016.

[5] S. Gupta and S. K. Jain, "A literature review of lean manufacturing," *International Journal of Management Science and Engineering Management*, vol. 8, no. 4, pp. 241–249, 2013.

[6] R. Petrolo, V. Loscri, and N. Mitton, "Cyber-Physical Objects as Key Elements for a Smart Cyber-City," in *Management of Cyber Physical Objects in the Future Internet of Things*, Internet of Things, pp. 31–49, Springer International Publishing, 2016.

[7] S. Ali, S. B. Qaisar, H. Saeed, M. F. Khan, M. Naeem, and A. Anpalagan, "Network challenges for cyber physical systems with tiny wireless devices: A case study on reliable pipeline condition monitoring," *Sensors (Switzerland)*, vol. 15, no. 4, pp. 7172–7205, 2015.

[8] A. Ojamaa, V. Kotkas, M. Spichakova, and J. Penjam, "Developing a lean mass customization based manufacturing," in *Proceedings of the 2013 16th IEEE International Conference on Computational Science and Engineering, CSE 2013*, pp. 28–33, Sydney, Australia, December 2013.

[9] H. Yetiş and M. Karaköse, "Adaptive vision based condition monitoring and fault detection method for multi robots at production lines in industrial systems," *International Journal of Applied Mathematics, Electronics and Computers*, vol. 4, no. 1, pp. 271-271, 2016.

[10] R. Bączyk and A. Kasiński, "Visual simultaneous localisation and map-building supported by structured landmarks," *International Journal of Applied Mathematics and Computer Science*, vol. 20, no. 2, pp. 281–293, 2010.

[11] G. Vegah, U. Wajid, and B. Adebisi, "Smart-agent system for flexible, personalised transport service," *The Journal of Engineering*, pp. 1–11, 2016.

[12] J. Kolodziej and F. Xhafa, "Modern approaches to modeling user requirements on resource and task allocation in hierarchical computational grids," *International Journal of Applied Mathematics and Computer Science*, vol. 21, no. 2, pp. 243–257, 2011.

[13] Krebson Security, "Hacked Cameras, DVRs Powered Today's Massive Internet Outage," https://krebsonsecurity.com/2016/10/hacked-cameras-dvrs-powered-todays-massive-internet-outage/, Oct. 2016.

[14] M. Hall-May, M. Surridge, and R. Nossal-Tüyeni, "Resilient critical infrastructure management with a service oriented architecture: A test case using airport collaborative decision making," *International Journal of Applied Mathematics and Computer Science*, vol. 21, no. 2, pp. 259–274, 2011.

[15] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDOS) flooding attacks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 2046–2069, 2013.

[16] G. Sen Gupta, M. Ooi, S. Khan et al., "Assistive technology for relieving communication lumber between hearing/speech impaired and hearing people," *The Journal of Engineering*, pp. 1–12, 2014.

[17] P. Coletti and T. Aichner, *Mass Customization*, Springer Berlin Heidelberg, Berlin, Germany, 2011.

[18] R. G. Hollands, "Will the real smart city please stand up?" *City*, vol. 12, no. 3, pp. 303–320, 2008.

[19] A. Coe, G. Paquet, and J. Roy, "E-governance and smart communities: a social learning challenge," *Social Science Computer Review*, vol. 19, no. 1, pp. 80–93, 2001.

[20] M. Karakose, "Reinforcement learning based artificial immune classifier," *The Scientific World Journal*, vol. 2013, Article ID 581846, 7 pages, 2013.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 4180801, 11 pages
https://doi.org/10.1155/2017/4180801

Research Article

Distributed Group-Based Mobility Management Scheme in Wireless Body Area Networks

Moneeb Gohar,[1] Hind Ahmed M. Alrubaish,[2] Ruba Suliman M. Alowaid,[3] and Jin-Ghoo Choi[4]

[1]*Department of Computer Science, Bahria University, Islamabad, Pakistan*
[2]*Department of Computer, Deanship of Preparatory Year, University of Dammam, Dammam, Saudi Arabia*
[3]*University of Dammam, Dammam, Saudi Arabia*
[4]*Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, Republic of Korea*

Correspondence should be addressed to Moneeb Gohar; moneebgohar@gmail.com and Jin-Ghoo Choi; jchoi@yu.ac.kr

Received 2 February 2017; Revised 19 April 2017; Accepted 26 April 2017; Published 23 May 2017

Academic Editor: Jaime Lloret

For group-based mobility management in 6LoWPAN-based wireless body area networks (WBAN), some schemes using the Proxy Mobile IPv6 (PMIP) have been proposed. However, the existing PMIP-based mobility schemes tend to induce large registration delay and handover delay. To overcome such limitations, we propose a new distributed group-based mobility management scheme, in which the Local Mobility Anchor (LMA) function is implemented by each Mobile Access Gateway (MAG) and the handover operation is performed between two neighboring MAGs without the help of LMA. Besides, each MAG maintains the information of the group of mobile sensors and aggregates the Authentication-Authorization-Accounting (AAA) query messages for a group of mobile sensors as a "single" message to decrease the control overhead. By numerical analysis, it is shown that the proposed scheme can reduce the registration and handover delays, compared to the existing PMIP-based mobility schemes.

1. Introduction

World population growth is facing challenges like increase of life expectancy leading to aging population and rise in healthcare costs. These challenges have triggered the introduction of novel technology-driven enhancements to current healthcare practices such as small and intelligent medical sensors which can be worn on or implanted in the human body. These sensors send their data to an external medical server. With these sensors, the patient experiences greater physical mobility. On the other hand, millions of people die from different fatal diseases like cancer and cardiovascular disease every year. This is because many people are diagnosed when it is too late. However, it is possible to utilize the latest technological advances in wireless body area network (WBAN) systems for the early detection and prevention of potential diseases that may occur later in people's lives. This can be done by using the sensors nodes on WBAN for continuous monitoring of health conditions [1–3].

In WBAN, sensors can be attached to the human body or clothes [4, 5]. These sensors can be used to measure the parameters associated with the human body. The measured values can be gathered and transmitted to the main server healthcare applications.

The IP shortage problem can be solved through IPv6 addresses, because it has large address spaces. Recently, Low-power Wireless Personal Area Networks (LoWPANs) have attracted lots of attention because of the Internet of Things (IOTs). The Internet Engineering Task Force (IETF) has set up a working group recently for IPv6 over Low-power Wireless Personal Area Networks (6LoWPANs) [6], in which it is possible to connect wireless sensor nodes to IPv6 networks. However, the wireless sensors nodes are unable to hold the complete IPv6 address because the maximum

packet size of IEEE 802.15.4 is 127 bytes. For this purpose, the 6LoWPAN adds an adaptation layer between MAC and the network layer to compress the 40-byte IPv6 packet header into 2 bytes for incoming communication and decompress it for the outgoing interface.

In mobile environments, it is required to provide sensor nodes with mobility management, such as handover control. The host-based approaches such as Mobile IPv6 (MIPv6) [7], Hierarchal Mobile IPv6 (HMIPv6) [8], and Fast Mobile IPv6 (FMIPv6) [9] are unsuitable to be applied in Internet Protocol (IP) based wireless sensor networks. All of the sensors have mobility stack and have actively participated in mobility related signaling. The network-based approaches, such as Proxy Mobile IPv6 (PMIPv6), can be considered as the most suitable approach for the mobility management in 6LoWPAN based WBAN.

Recently, several schemes have been proposed to support mobility in the 6LoWPAN-based WBAN [10, 11], which usually use the Proxy Mobile IPv6 (PMIP) protocol [12]. These schemes reduce the registration and handover delays, but there remains much room for improvement in the delay performance.

In this paper, we propose an enhanced group-based mobility scheme for 6LoWPAN-based WBAN by using the distributed PMIP approach so as to further reduce registration and handover delays. In the proposed scheme, the Local Mobility Anchor (LMA) function is implemented by each Mobile Access Gateway (MAG), and MAGs perform the handover control operations for the group of sensors. The MAGs also perform the aggregation of Authentication-Authorization-Accounting (AAA) query messages incurred by the movement of the group of sensor nodes. Different from the standard PMIP protocol, each MAG does not perform the binding update and deregistration operations with LMA, since it has the LMA functionality.

The rest of this paper is organized as follows. Section 2 describes the related works and existing mobility schemes for comparison and motivation. Section 3 describes the proposed distributed group-based mobility scheme. Section 4 demonstrates the performance analysis by comparing the proposed and existing schemes by numerical analysis. Section 5 presents the numerical results and discussion. Section 6 concludes this paper.

2. Related Works

In this section, first related works are described and then the existing group-based mobility schemes are considered for comparison in this paper.

2.1. Related Works. Reference [13] provides a survey of sensor devices and protocols on physical layer, data link layer, and radio technology aspects of body area networks. This paper also highlighted some of the design challenges and open issues on body area networks that still need to be addressed. In [14], it is shown that IPv6 over Low-power Wireless Personal Area Networks (6LoWPANs) is suitable for wireless body sensor networks because the sensors are based on IEEE 802.15.4 standard. The 6LoWPAN mobility has been an important solution in wireless body sensor networks. Most of the existing host-based mobility [7–9, 15] protocols for IPv6 are unsuitable for 6LoWPAN based wireless sensor networks, because they are tunnel based approaches, which shows that all of the mobile sensors actively participated in mobility related signaling in order to ensure the continuity of communications. The network-based approach [12, 16] is suitable to handle the mobility management of 6LoWPAN based body sensors. On behalf of sensors, the network node exchanges the mobility related signaling and there is no need to perform Duplicate Address Detection (DAD) of the IP address and thus it reduces the control signaling cost and handover latency of each sensor. In [17], the author highlighted the requirements and resources for adapting the existing solutions to 6LoWPAN based on the notion that single mobile node's perspective is still a challenge and further researches are required. Therefore, [18] proposed a scheme to support Internet Protocol (IP) wireless sensor networks in which the packet format in ingress interface is to deal with the handoff procedure. In [17], it is also shown that multihop communications are not supported by Proxy Mobile IP (PMIP) protocol. Hence, in [19], a multihop communication scheme has been proposed in the ingress interface by using mesh routing. But this scheme is only on single mobile node perspective. So, if there is a group of mobile nodes moving around together, then this scheme [19] has no effective solution. In [20], a 6LoWPAN based mobility scheme has been proposed which depends on the dispatch types. But this scheme does not have a clear improvement on reducing handover delay. From the network-based perspective, in [21], a Network Mobility (NEMO) [10] based 6LoWPAN mobility scheme has been proposed, in which a mobile router performs the handoff and modifies the 6LoWPAN dispatch. However, this protocol overloaded the mobile router. To maintain the sessions while reducing the handover delay and signaling cost of group-based mobility in 6LoWPAN-based wireless sensor body area networks is an important issue. Therefore, [22] considered lots of correlated sensor nodes moving together and performed handoff together at the same time in PMIPv6 environment. The Local Mobility Anchor (LMA) calculates the SNR value for each sensor and makes a group of sensors with similar SNR values. This protocol sends Proxy Binding Ack (PBA) per group which can reduce the handoff signaling cost and save some additional unnecessary handoff messages such as Proxy Binding Update (PBU) and deregistration PBU (DeReg PBU). However, this protocol does not reduce the Router Solicitation (RS) and Router Advertisement (RA) messages in case body sensors move in the PMIPv6 domain together. Therefore, this protocol is not suitable for the 6LoWPAN-based wireless sensor body network. So, in [23], network architecture is proposed which integrates NEMO and 6LoWPAN. The mechanism of group mobility management and the signaling process containing the registration at the home network, association negotiation, handoff between different access routers, and packet routing are discussed. However, this protocol increases the burden of the sensor router due to connection and support for

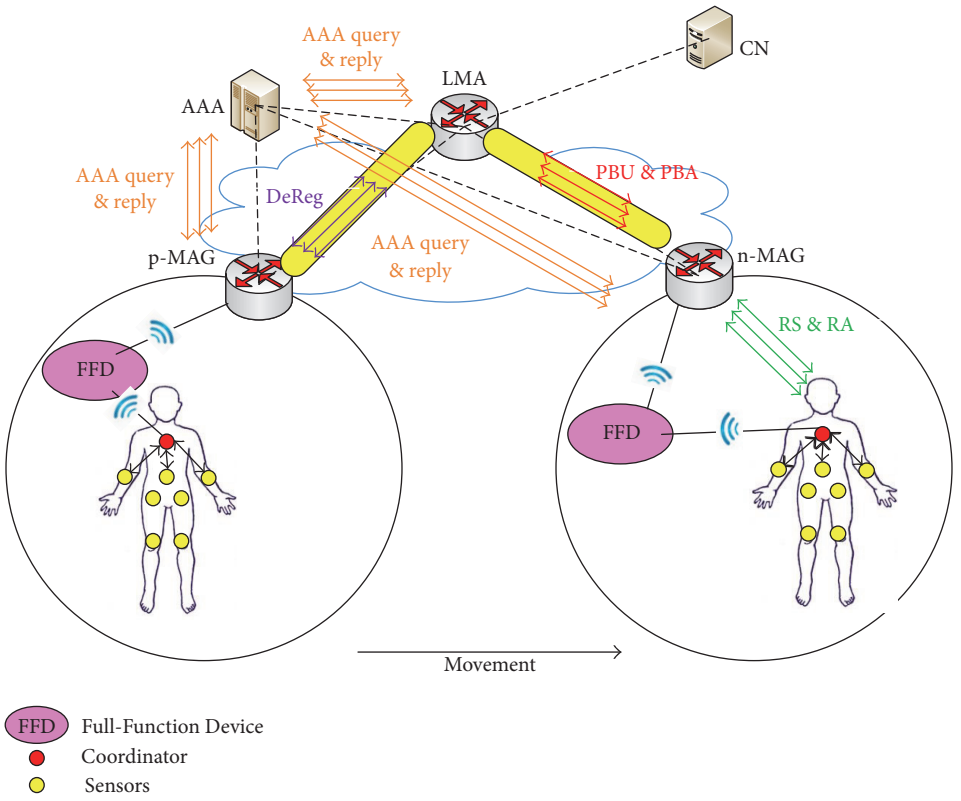


Figure 1: Existing PMIP mobility schemes.

other normal sensor nodes inside the group. This protocol is designed for MIPv6, not for PMIPv6.

2.2. Existing Group-Based Mobility Schemes for Comparison

2.2.1. PMIP [12]. In this scheme, the standard PMIP is used for the group-based mobility support [11]. As shown in Figure 1, when a set of body sensors are detached from a previous MAG (p-MAG), the p-MAG sends *deregistration (DeReg)* messages to LMA for each body sensor. When the body sensors are attached to a new MAG (n-MAG), each body sensor sends a *Router Solicitation (RS)* message to n-MAG via the *coordinator* and Full-Function Device (FFD) which are preconfigured in the 6LoWPAN network [6].

Upon reception of RS messages from body sensors, n-MAG will send the *Authentication-Authorization-Accounting (AAA) query* messages for authentication for all body sensors. After authentication, AAA server responds with *AAA reply* messages, containing the LMA address, to n-MAG. Then, n-MAG will send *Proxy Binding Update (PBU)* messages to LMA for each body sensor. Now, LMA will perform the AAA query operation with AAA server by exchanging *AAA query* and *reply* messages for each sensor. After that, LMA sends *Proxy Binding Ack (PBA)* messages to n-MAG in response to each respective *PBU* message. Finally, n-MAG responds with *Router Advertisement (RA)* messages to the body sensors.

2.2.2. PMIP-Group [10]. The standard PMIP-based scheme has a drawback that a lot of *DeReg*, *PBU*, and *PBA* messages are exchanged between LMA and MAGs (p-MAG and n-MAG) for all body sensors.

To enhance this PMIP scheme [12], the PMIP-Group [10] was proposed, in which a single *DeReg* message is exchanged between p-MAG and LMA by aggregating the associated messages from all sensors, as shown in Figure 2. In case of handover to n-MAG, the *PBU* and *PBA* messages are also aggregated between LMA and n-MAG. By this aggregation, the number of *PBU* and *PBA* messages can be reduced between LMA and n-MAG.

2.2.3. PMIP-Coordinator [11]. The PMIP-Group [7] scheme tends to exchange many *RS/RA* messages between sensors and MAGs (p-MAG and n-MAG). The PMIP-Coordinator scheme [11] was proposed for further enhancement of PMIP-Group scheme [10]. In this scheme, the *coordinator* will communicate with n-MAG on behalf of the body sensors, as shown in Figure 3.

When the *coordinator* is attached to n-MAG, then it sends a single *Router Solicitation (RS)* message, which contains the associated group information, MN-IDs, and link-layer address, to n-MAG, by way of FFD at a time. Then, n-MAG responds with a *Router Advertisement (RA)* message to the coordinator in response to the *RS* message.

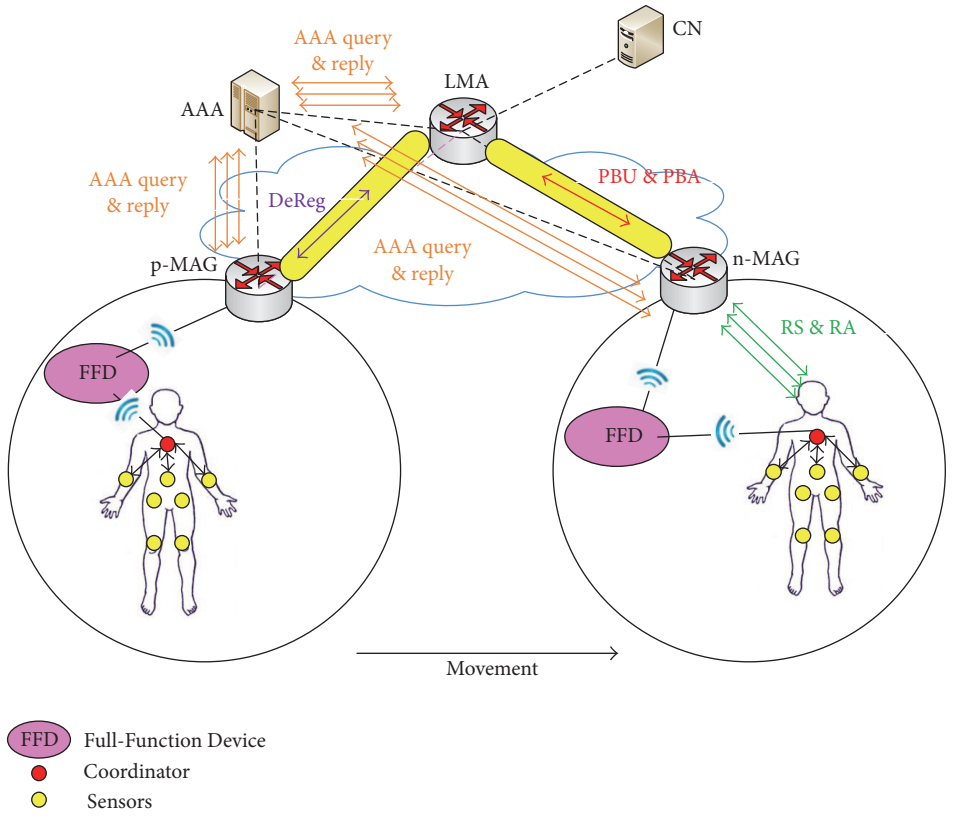


Figure 2: Existing PMIP-Group mobility schemes.

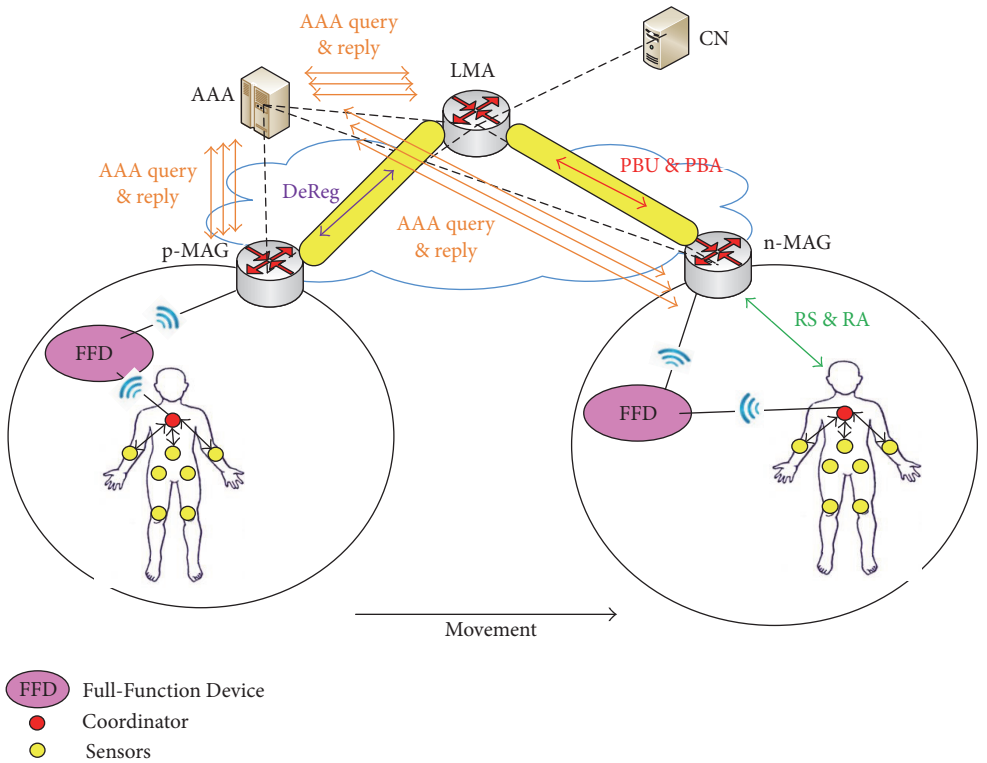


Figure 3: Existing PMIP-Coordinator mobility schemes.

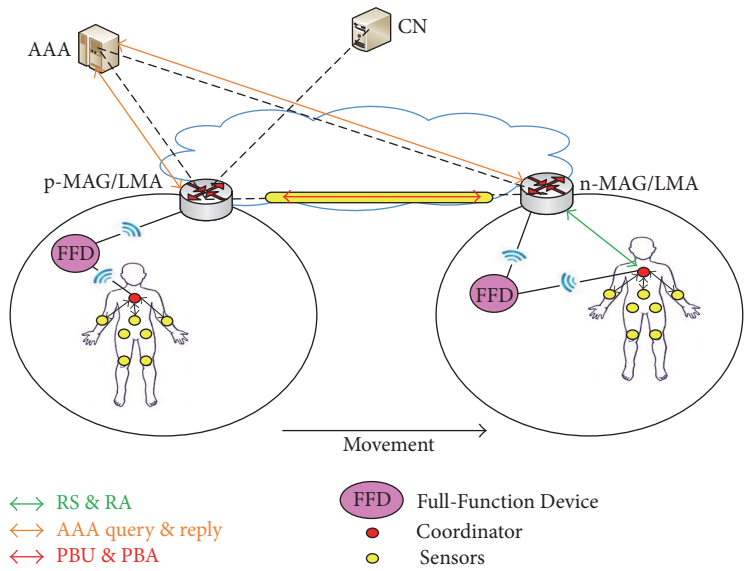


Figure 4: Network model for the proposed scheme.

3. Distributed Group-Based Mobility Management Scheme

This section first introduces the network model of the proposed scheme and then describes the initial registration and handover operations.

3.1. Network Model. Figure 4 shows the network model for the proposed scheme. We consider a group of IPv6 over Low-power Wireless Personal Area Network (6LoWPAN) sensors attached to the human body. One of them acts as a coordinator, and only the coordinator can exchange the control signaling messages with the Mobile Access Gateway (MAG) on behalf of the other sensors. The 6LoWPAN domain contains a Full-Function Device (FFD). In the proposed scheme, the Local Mobility Anchor (LMA) function is implemented by each MAG, and the handover operation is performed between two neighboring MAGs without the help of LMA for the group of sensors. Each MAG maintains the information of the group of sensors and aggregates the *Authentication-Authorization-Accounting (AAA)* query messages to the AAA server in the authentication process. Each MAG does not perform the binding update and deregistration operation with LMA, since it has the LMA functionality.

Initially, the coordinator communicates with the correspondent node (CN) in the previous Mobile Access Gateway (p-MAG) domain, and then it moves to a new Mobile Access Gateway (n-MAG) by handover.

3.2. Initial Registration. The aim of the registration phase is to reduce the amount of control messages. Figure 5 illustrates the initial registration of the proposed scheme. When a group of body sensors enter a PMIP domain and the coordinator is attached to MAG/LMA, it sends an aggregated *Router Solicitation (RS)* message, containing the information on group, MN-IDs, and Link-Layer Addresses (LLAs), to MAG/LMA (Step 1).

Upon reception of this *RS* message from the coordinator, MAG/LMA sends an aggregated *AAA query* message with all LLAs for authentication to the AAA server. The aggregated *AAA query* message is different from the previous schemes, in that it contains LLAs of a group of sensors.

After authentication, the AAA server responds with a "single" *AAA reply* message to MAG/LMA (Steps 2 and 3). After that, MAG initiates the DHCP solicitation procedure to request home network prefixes (HNPs) for the body sensors one by one. Then, the DHCP server replies with the unique HNPs for each body sensor. The MAG creates the binding cache entry; store the home network prefix information. The DHCP server also configures the respective Home-of-Address (HoA) from those prefixes and sends it to MAG. Then, the MAG/LMA responds to the coordinator with an RA message as a response to the *RS* message (Step 4).

3.3. Handover Operations. The body sensors perform active scans periodically with nearby FFDs by sending a beacon request. The nearby FFDs receive the beacon request from the body sensors and advertise the beacon message containing their MAG-IDs to the body sensors. The body sensors receive the beacon message and decide whether it still resides in the same MAG or has moved to another MAG by comparing the current MAG-IDs with the previous MAG-IDs. If the current MAG-ID and previous MAG-ID are the same, then the movement represents intra-PAN mobility. On the other hand, if the current MAG-ID is different from the previous MAG-ID, then the body sensors are able to detect its movement from the old MAG to the new MAG.

Figure 6 shows the detailed signaling flow for the handover procedure. We now assume that the coordinator changes its point of attachment in the same network domain. When the coordinator is detached from p-MAG/LMA and

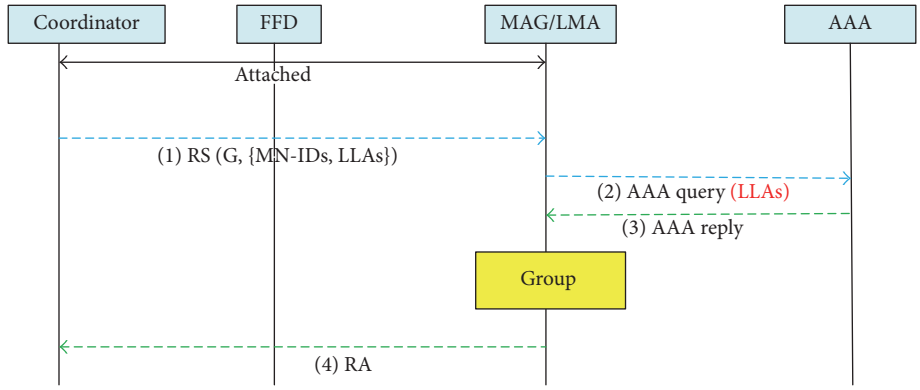


Figure 5: Initial registration.

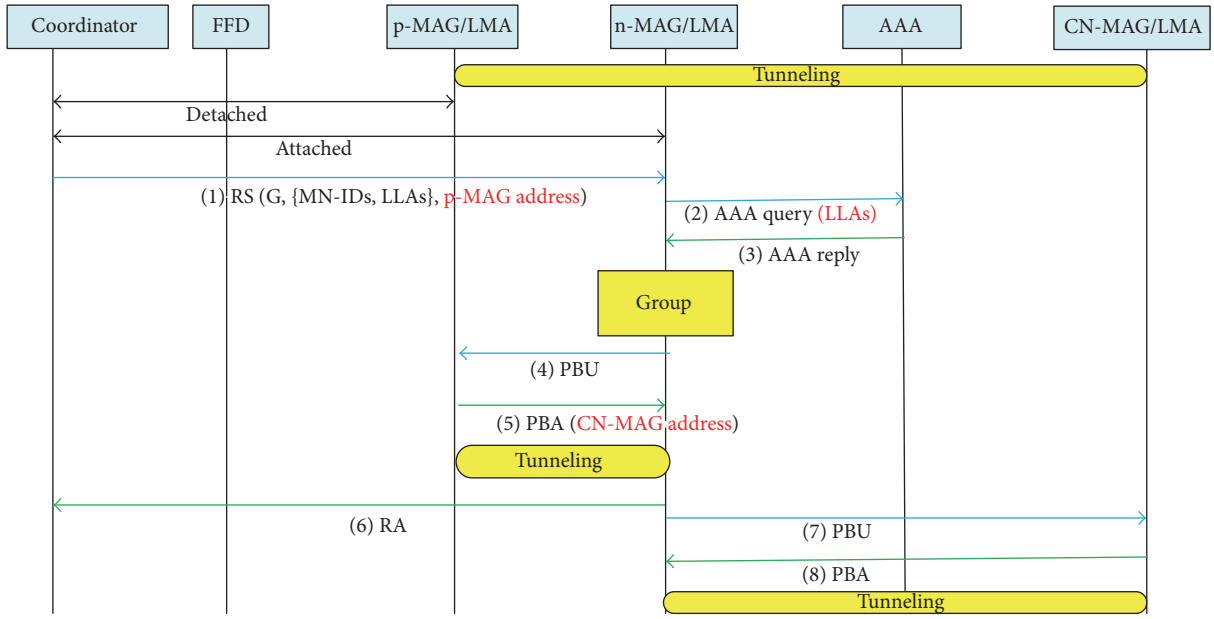


Figure 6: Handover operations.

attached to n-MAG/LMA, then the coordinator sends an aggregated *RS* message to n-MAG/LMA (Step 1). After that, n-MAG/LMA will exchange the aggregated *AAA query* and *reply* messages with AAA server (Steps 2 and 3). After authentication, n-MAG sends a *Proxy Binding Update (PBU)* message to p-MAG/LMA for handover control (Step 4). Now, n-MAG/LMA obtains the location of p-MAG/LMA from the *RS* message. After that, the p-MAG/LMA responds with *Proxy Binding Ack (PBA)* to n-MAG/LMA (Step 5). The *PBA* message shall include the information of CN-MAG/LMA address, which is recorded into the mapping table of p-MAG/LMA. So, the handover tunnel is established between p-MAG/LMA and n-MAG/LMA for data forwarding. After establishment of handover tunnel, n-MAG responds to the coordinator with a *Router Advertisement (RA)* message (Step 6).

Now, n-MAG/LMA sends a *PBU* message to CN-MAG/LMA for route optimization (Step 7). On reception of the *PBU* message, CN-MAG/LMA will update its mapping table and send a *PBA* to n-MAG/LMA (Step 8). n-MAG/LMA and CN-MAG/LMA will now use the optimized route.

4. Performance Analysis

For performance analysis, we compare the registration and handover delays for the four candidate mobility schemes: PMIP, PMIP-Group, PMIP-Coordinator, and the proposed scheme.

4.1. Analysis Model. We consider a network illustrated in Figure 7, in which each wired/wireless link is represented by bandwidth, latency, and average queuing delay. We adopt a generic model for Multiple Access Control (MAC) scheme to focus on the analysis of registration delay and handover delay associated with the proposed mobility scheme.

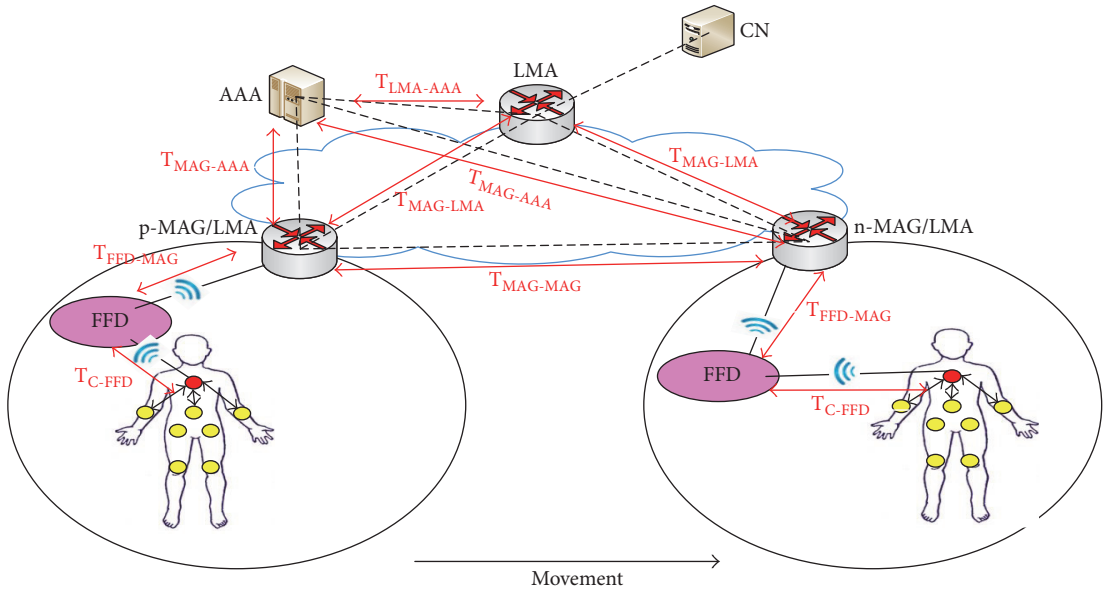


Figure 7: Network model for performance analysis.

We summarize the notations used in our analysis in the Parameters Used for Analysis.

In Figure 7, we denote by $T_{x\text{-}y}(S)$ the transmission delay of a message with size S sent from x to y via the "*wireless*" link, where each message can experience the failure at the probability of q by using "*iid*" error model. Then, $T_{x\text{-}y}(S)$ can be expressed as $T_{x\text{-}y}(S) = [1/(1-q)] \times [(S/B_{wl}) + L_{wl}]$. In the meantime, we denote by $T_{x\text{-}y}(S, H_{x\text{-}y})$ the transmission delay of a message with size S sent from x to y via the "*wired*" link, where $H_{x\text{-}y}$ represents the number of wired hops between node x and node y. Then, $T_{x\text{-}y}(S, H_{x\text{-}y})$ is expressed as $T_{x\text{-}y}(S, H_{x\text{-}y}) = H_{x\text{-}y} \times [(S/B_w) + L_w + T_q]$.

4.2. Registration Delay (RD). As shown in Figure 1, when the body sensors are attached to a MAG, they send *RS* messages to MAG via the coordinator and FFD. After that, MAG performs the AAA operations for authentication for all body sensors. Then, MAG performs the *PBU* operation with LMA for all body sensors. Now, LMA performs the AAA operations with AAA server for all body sensors. After authentication, LMA responds with the *PBA* message to MAG. Now, MAG will respond with the *RA* messages to all of the body sensors. Accordingly, we get the RD of PMIP as follows:

$$\begin{aligned} RD_{PMIP} = N_S \times \{&2T_{C\text{-}FFD}(S_c) + 2T_{FFD\text{-}MAG}(S_c) \\ &+ 2T_{MAG\text{-}AAA}(S_c) + 2T_{MAG\text{-}LMA}(S_c) \\ &+ 2T_{LMA\text{-}AAA}(S_c)\}. \end{aligned} \quad (1)$$

As shown in Figure 2, the whole operations of PMIP-Group are the same as those of PMIP. Different from PMIP, the PMIP-Group scheme uses the aggregated *PBU* and *PBA* messages between LMA and MAGs. Accordingly, we get the RD of PMIP-Group as follows:

$$\begin{aligned} RD_{PMIP\text{-}Group} = N_S \times \{&2T_{C\text{-}FFD}(S_c) + 2T_{FFD\text{-}MAG}(S_c) \\ &+ 2T_{MAG\text{-}AAA}(S_c) + 2T_{LMA\text{-}AAA}(S_c)\} \\ &+ 2T_{MAG\text{-}LMA}(S_c). \end{aligned} \quad (2)$$

As shown in Figure 3, the PMIP-Coordinator uses the aggregated *RS* and *RA* messages between MAG and the coordinator. Thus, we get the RD of PMIP-Coordinator as

$$\begin{aligned} &RD_{PMIP\text{-}Coordinator} \\ &= 2T_{C\text{-}FFD}(S_c) + 2T_{FFD\text{-}MAG}(S_c) + 2T_{MAG\text{-}LMA}(S_c) \\ &\quad + N_S \times \{2T_{MAG\text{-}AAA}(S_c) + 2T_{LMA\text{-}AAA}(S_c)\}. \end{aligned} \quad (3)$$

In the proposed scheme, when the coordinator is attached to n-MAG, then it sends an aggregated *RS* message on behalf of all body sensors to n-MAG by way of FFD. After that, n-MAG performs the AAA operations with AAA server. Based on this, n-MAG performs the authentication process and responds with an aggregated *RA* message to the coordinator. Thus, we get the RD of the proposed scheme as follows:

$$\begin{aligned} RD_{Proposed\ Scheme} = &2T_{C\text{-}FFD}(S_c) + 2T_{FFD\text{-}MAG}(S_c) \\ &+ 2T_{MAG\text{-}AAA}(S_c). \end{aligned} \quad (4)$$

4.3. Handover Delay (HD). The handover delay is defined as the gap between the time that body sensors cannot receive the packets from p-MAG and the time that body sensors receive the first packet from n-MAG.

As shown in Figure 1, when the body sensors are detached from p-MAG, then p-MAG will send *DeReg* messages to LMA for body sensors. When the body sensors are attached to n-MAG, they send *RS* messages to n-MAG by way of the coordinator and FFD. After that, n-MAG performs the AAA query and reply operations with AAA server for authentication for all body sensors, and then n-MAG performs the *PBU* operations with LMA for all body sensors. Now, LMA performs the AAA query/reply operations with AAA server for all body sensors. After authentication, LMA responds with the *PBA* message to n-MAG. The handover tunnel will be established between n-MAG and LMA. After the tunnel establishment, n-MAG responds with the *RA* messages to all body sensors. Accordingly, we get the HD of PMIP as follows:

$$\begin{aligned} \mathrm{HD}_{\mathrm{PMIP}} &= N_S \times \{2T_{\text{C-FFD}}(S_c) + 2T_{\text{FFD-MAG}}(S_c) \\ &\quad + 2T_{\text{MAG-AAA}}(S_c) + 4T_{\text{MAG-LMA}}(S_c) \\ &\quad + 2T_{\text{LMA-AAA}}(S_c)\} + T_{\text{MAG-LMA}}(S_d). \end{aligned} \tag{5}$$

As shown in Figure 2, the PMIP-Group scheme uses the aggregated *DeReg* and *PBU/PBA* messages. Accordingly, we get the HD of PMIP-Group as follows:

$$\begin{aligned} \mathrm{HD}_{\text{PMIP-Group}} &= N_S \times \{2T_{\text{C-FFD}}(S_c) + 2T_{\text{FFD-MAG}}(S_c) \\ &\quad + 2T_{\text{MAG-AAA}}(S_c) + 2T_{\text{LMA-AAA}}(S_c)\} \\ &\quad + 4T_{\text{MAG-LMA}}(S_c) + T_{\text{MAG-LMA}}(S_d). \end{aligned} \tag{6}$$

As shown in Figure 3, the PMIP-Coordinator scheme uses the aggregated *RS* and *RA* messages between MAG and the coordinator. Accordingly, we can get the HD of PMIP-Coordinator as follows:

$$\begin{aligned} &\mathrm{HD}_{\text{PMIP-Coordinator}} \\ &\quad = 2T_{\text{C-FFD}}(S_c) + 2T_{\text{FFD-MAG}}(S_c) + 4T_{\text{MAG-LMA}}(S_c) \\ &\qquad + N_S \times \{2T_{\text{MAG-AAA}}(S_c) + 2T_{\text{LMA-AAA}}(S_c)\} \\ &\qquad + T_{\text{MAG-LMA}}(S_d). \end{aligned} \tag{7}$$

In the proposed scheme, when the coordinator is attached to n-MAG, it sends an aggregated *RS* message on behalf of body sensors to n-MAG by way of FFD. After that, n-MAG performs the AAA operations with the AAA server by aggregation. Then, n-MAG exchanges the *PBU* and *PBA* messages with p-MAG for establishment of a handover tunnel. After tunnel establishment, n-MAG responds with an aggregated *RA* message to the coordinator. Accordingly, we get the HD of the proposed scheme as follows:

$$\begin{aligned} \mathrm{HD}_{\text{Proposed Scheme}} &= 2T_{\text{C-FFD}}(S_c) + 2T_{\text{FFD-MAG}}(S_c) \\ &\quad + 2T_{\text{MAG-AAA}}(S_c) \\ &\quad + 2T_{\text{MAG-MAG}}(S_c) \\ &\quad + T_{\text{MAG-MAG}}(S_d). \end{aligned} \tag{8}$$

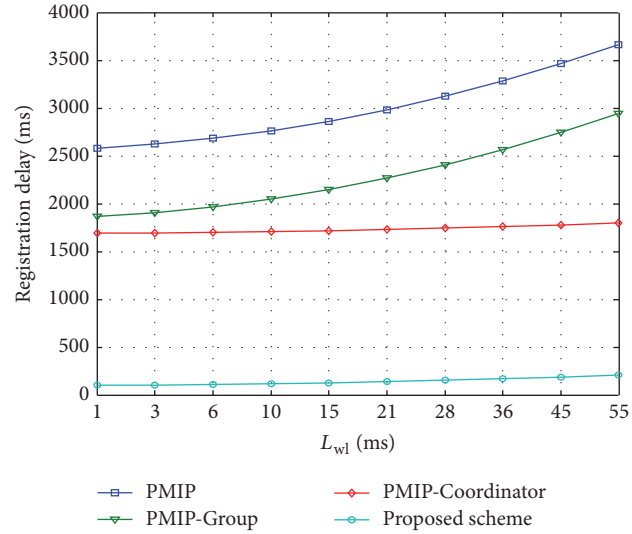


Figure 8: Impact of L_{wl} on registration delay.

5. Numerical Results and Discussion

Based on the analysis given so far, we now compare the performances of the candidate schemes.

For numerical analysis, by referring to [24], the default value of each parameter is configured as follows: $H_{\text{MAG-LMA}} = H_{\text{MAG-MAG}} = 5$, $H_{\text{MAG-AAA}} = H_{\text{LMA-AAA}} = 5$, $L_{\text{wl}} = 10$ ms, $L_w = 2$ ms, $q = 0.5$, $T_q = 5$ ms, $N_S = 10$, $S_c = 96$ bytes, $S_d = 200$ bytes, $B_{\text{wl}} = 11$ Mbps, and $B_w = 100$ Mbps. Among these parameters, we note that L_{wl}, T_q, and N_s may depend on the network conditions. Thus, we will compare the performances of candidate schemes by varying those parameter values.

5.1. Registration Delay. Figure 8 shows the impact of wireless link delay (L_{wl}) on registration delay. From the figure, we can see that the registration delay linearly increases as L_{wl} gets larger, for all candidate schemes. It is shown that the proposed scheme gives better performance than the existing schemes. This is because the proposed scheme uses the LMA functionality over MAG, and thus the PBU and PBA operations are not required between LMA and MAG. In the proposed scheme, the MAG also performs aggregated AAA query and reply operations.

Figure 9 compares the registration delay for different average queuing delay (T_q) at each node. It is shown in the figure that the registration delay linearly increases as T_q gets larger, for all candidate schemes. We can see that the proposed scheme gives the best performance among the candidate schemes. This is because each MAG implements the LMA functionality.

Figure 10 illustrates the impact of the number of sensors (N_s) on registration delay. In the figure, the handover delay linearly increases as N_s gets larger, for the existing schemes. It is shown that the proposed scheme is not affected by N_s. This is because the amounts of the AAA query and reply messages for each registration are constant no matter what the number of the body sensors is. Besides, the proposed scheme also uses

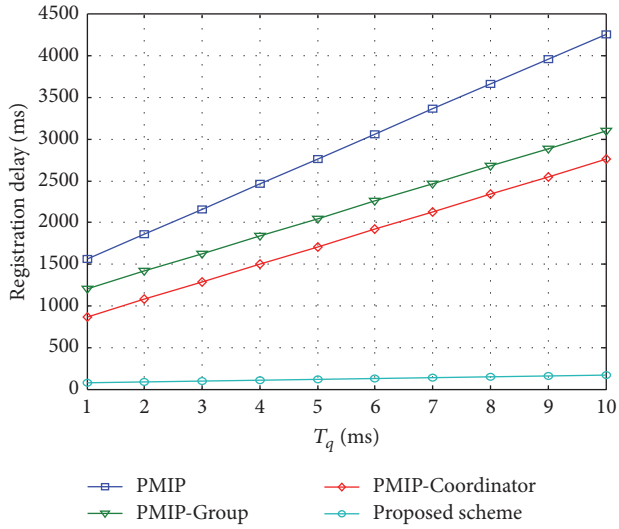


Figure 9: Impact of T_q on registration delay.

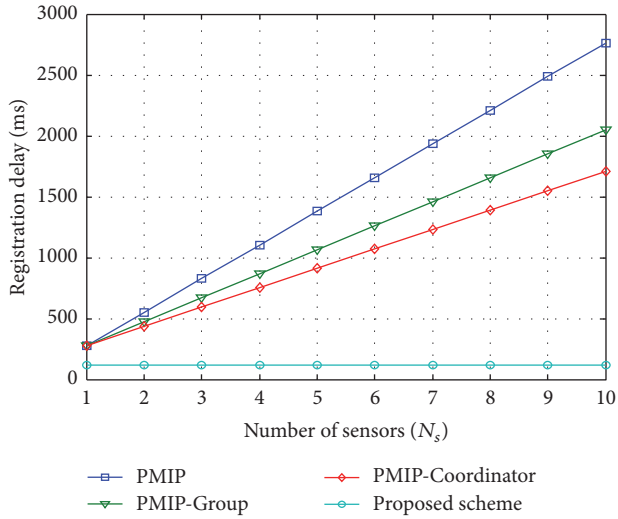


Figure 10: Impact of N_s on registration delay.

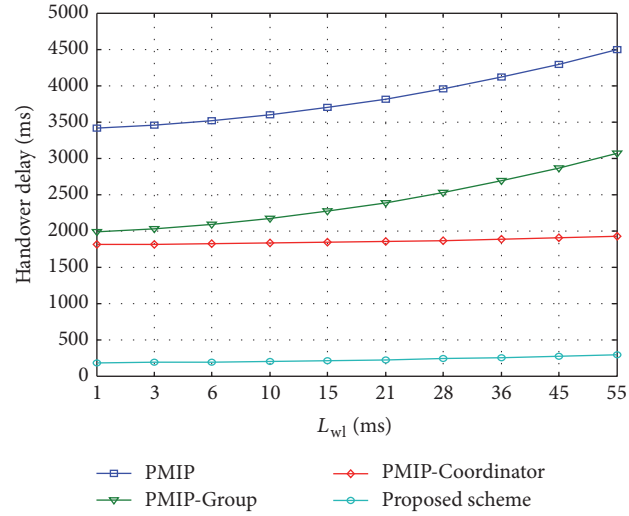


Figure 11: Impact of L_{wl} on handover delay.

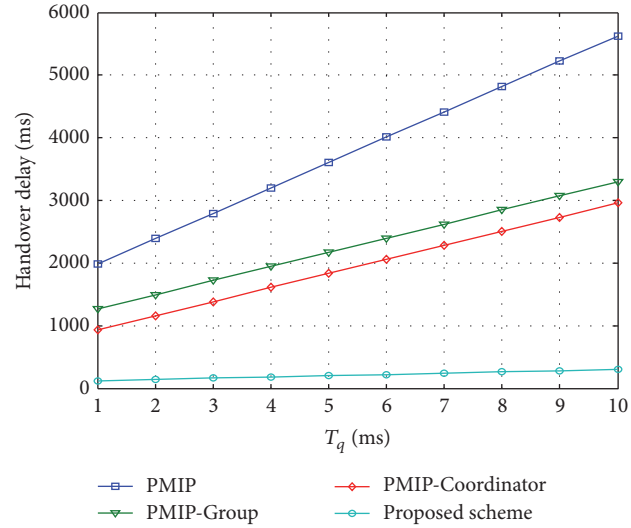


Figure 12: Impact of T_q on handover delay.

the LMA functionality over MAG, and thus the PBU and PBA operations are not required between LMA and MAG. Hence, each body sensor needs to transmit a smaller amount of control messages and incurs lower registration delay as the number of body sensors increases.

5.2. Handover Delay. Figure 11 illustrates the impact of wireless link delay (L_{wl}) on handover delay. In the figure, the handover delay linearly increases as L_{wl} gets larger, for all candidate schemes. It is shown that the proposed scheme gives better performance than the existing schemes. This is because the LMA function is implemented by each MAG and the handover operation is performed between two neighboring MAGs without the help of LMA.

Figure 12 shows the impact of average queuing delay (T_q) on handover delay. It is shown in the figure that the handover delay linearly increases as T_q gets larger, for all the candidate schemes. We can see that the proposed scheme gives the best performance among the candidate schemes.

Figure 13 shows the impact of the number of sensors (N_s) on handover delay. In the figure, the handover delay linearly increases as N_s gets larger, for the existing schemes. It is shown that the proposed scheme is not affected by N_s. This is because the amounts of the AAA query and reply messages for each handoff are constant no matter what the number of the body sensors is. Besides, the proposed scheme also implemented the LMA functionality over MAG, and the handover operation is performed between two neighboring MAGs without the help of LMA and thus the deregistration operations are not required between LMA and p-MAG. Hence, each body sensor needs to transmit a smaller amount of control messages and incurs lower handover delay as the number of body sensors increases.

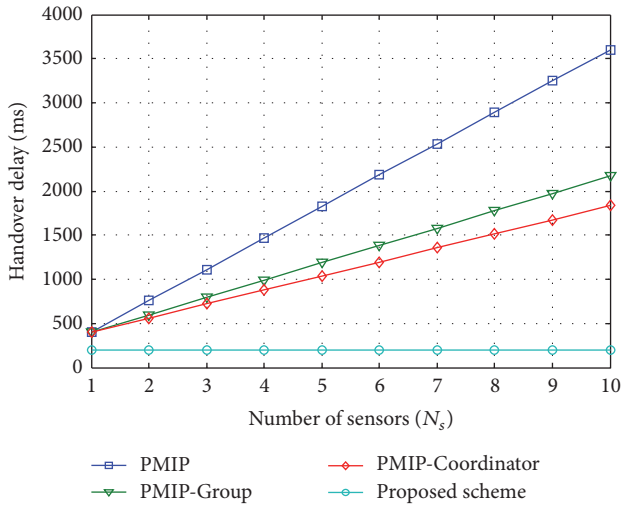


FIGURE 13: Impact of N_s on handover delay.

6. Conclusions

In this paper, we proposed a distributed group-based mobility management scheme in 6LoWPAN-based wireless body area networks. In the proposed scheme, the LMA function is implemented by each MAG, and the handover operation is performed between the two neighboring MAGs. Furthermore, the MAG aggregates the AAA query messages for the group of moving sensors. From the numerical results, we see that the proposed scheme can provide better performance than the existing schemes in terms of registration and handover delays.

Parameters Used for Analysis

S_c: Size of control packets (bytes)
S_d: Size of data packets (bytes)
N_S: Number of sensors in the domain
B_w: Wired link bandwidth (Mbps)
B_{wl}: Wireless bandwidth (Mbps)
L_w: Wired link delay (ms)
L_{wl}: Wireless link delay (ms)
$H_{a\text{-}b}$: Hop count between nodes a and b in the network
q: Wireless link failure probability
T_q: Average queuing delay at each node.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the Deanship of Scientific Research, University of Dammam, Saudi Arabia (2016-410-PYSS). This work was also supported by the 2016 Yeungnam University Research Grant.

References

[1] M. Ghamari, B. Janko, R. S. Sherratt, W. Harwin, R. Piechockic, and C. Soltanpur, "A survey on wireless body area networks for ehealthcare systems in residential environments," *Sensors*, vol. 16, no. 6, pp. 831–864, 2016.

[2] R. Cavallari, F. Martelli, R. Rosini, C. Buratti, and R. Verdone, "A survey on wireless body area networks: technologies and design challenges," *IEEE Communications Surveys and Tutorials*, vol. 16, no. 3, pp. 1635–1657, 2014.

[3] R. Sharma, H. S. Ryait, and A. K. Gupta, "Wireless body area nework—a review," *Research Cell*, vol. 17, 2016.

[4] J. Xing and Y. Zhu, "A survey on body area network," in *Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '09)*, September 2009.

[5] M. Gohar, J.-G. Choi, and S.-J. Koh, "An ID/locator separation based group mobility management in wireless body area network," *Journal of Sensors*, vol. 2015, Article ID 537205, 12 pages, 2015.

[6] N. Kushalnagar, G. Montenegro, and C. Schumacher, "IPv6 over low-power wireless personal area networks (6LoWPANs): overview, assumptions, problem statement, and goals," RFC, 4919, IETF, 2007.

[7] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," RFC 3775, IETF, 2004.

[8] D. Saha, A. Mukherjee, I. S. Misra, and M. Chakraborty, "Mobility support in IP: a survey of related protocols," *IEEE Network*, vol. 18, no. 6, pp. 34–40, 2004.

[9] H. Soliman and C. Castelluccia, *Hierarchical Mobile IPv6 Mobility Management*, British Royal Flying Corp, 2005.

[10] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, "Network Mobility (NEMO) basic support protocol," RFC 3963, 2005.

[11] Y.-S. Chen, C.-S. Hsu, and H.-K. Lee, "An enhanced group mobility protocol for 6lowpan-based wireless body area networks," *IEEE Sensors Journal*, vol. 14, no. 3, pp. 797–807, 2014.

[12] S. Hyeon, Y.-H. Han, H.-B. Lee, and H.-Y. Choi, "Empirical performance evaluation of IETF mobile IPv6 and proxy mobile IPv6," Proceedings of the international conference on mobile technology,applications, and systems, mobility'08 Article 68, 2008.

[13] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: a survey," *Mobile Networks and Applications*, vol. 16, no. 2, pp. 171–193, 2011.

[14] R. S. H. Istepanian, E. Jovanov, and Y. T. Zhang, "Introduction to the special section on m-Health: beyond seamless mobility and global wireless health-care connectivity," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 4, pp. 405–414, 2004.

[15] G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, "Transmission of IPv6 packets over IEEE 802.15.4 networks," RFC 4944, 2007.

[16] M. Shin, T. Camilo, J. Silva, and D. Kaspar, "Internet-draft-6LoWPAN mobility," *Internet-Draft*, vol. 2, pp. 5–10, 2007.

[17] L. M. L. Oliveira, A. F. de Sousa, and J. J. P. C. Rodrigues, "Routing and mobility approaches in IPv6 over LoWPAN mesh networks," *International Journal of Communication Systems*, vol. 24, no. 11, pp. 1445–1466, 2011.

[18] M. M. Islam and E.-N. Huh, "Sensor proxy mobile IPv6 (SPMIPv6)—a novel scheme for mobility supported IP-WSNs," *Sensors*, vol. 11, no. 2, pp. 1865–1887, 2011.

[19] J. Kim, R. Haw, E. J. Cho, C. S. Hong, and S. Lee, "A 6LoWPAN sensor node mobility scheme based on proxy mobile IPv6," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 2060–2072, 2012.

[20] G. Bag, M. T. Raza, K.-H. Kim, and S.-W. Yoo, "LoWMob: intra-PAN mobility support schemes for 6LoWPAN," *Sensors*, vol. 9, no. 7, pp. 5844–5877, 2009.

[21] J. H. Kim, C. S. Hong, and T. Shon, "A lightweight NEMO protocol to support 6LoWPAN," *ETRI Journal*, vol. 30, no. 5, pp. 685–695, 2008.

[22] Y. Li, Y. Jiang, H. Su, D. Jin, L. Su, and L. Zeng, "A group-based handoff scheme for correlated mobile nodes in proxy mobile IPv6," in *Proceedings of IEEE Global Telecommunications Conference, GLOBECOM 2009*, pp. 1–6, December 2009.

[23] R. Chai, Y.-L. Zhao, Q.-B. Chen, T. Dong, and W.-G. Zhou, "Group mobility in 6LoWPAN-based WSN," in *Proceedings of International Conference on Wireless Communications and Signal Processing, WCSP 2010*, pp. 1–5, October 2010.

[24] C. Makaya and S. Pierre, "An analytical framework for performance evaluation of IPv6-based mobility management protocols," *IEEE Transactions on Wireless Communications*, vol. 7, no. 3, pp. 972–983, 2008.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 6474768, 7 pages
https://doi.org/10.1155/2017/6474768

Research Article

A Reinforcement Learning Approach to Access Management in Wireless Cellular Networks

Jihun Moon[1] and Yujin Lim[2]

[1]*Department of Computer Science, University of Suwon, San 2-2, Wau-ri, Bongdam-eup, Hwaseong, Gyeonggi-do 445-743, Republic of Korea*
[2]*Department of Information Technology Engineering, Sookmyung Women's University, 100 Cheongpa-ro 47-gil, Yongsan-gu, Seoul 04310, Republic of Korea*

Correspondence should be addressed to Yujin Lim; yujin91@sookmyung.ac.kr

Received 25 February 2017; Accepted 9 April 2017; Published 14 May 2017

Academic Editor: Syed Hassan Ahmed

In smart city applications, huge numbers of devices need to be connected in an autonomous manner. 3rd Generation Partnership Project (3GPP) specifies that Machine Type Communication (MTC) should be used to handle data transmission among a large number of devices. However, the data transmission rates are highly variable, and this brings about a congestion problem. To tackle this problem, the use of Access Class Barring (ACB) is recommended to restrict the number of access attempts allowed in data transmission by utilizing strategic parameters. In this paper, we model the problem of determining the strategic parameters with a reinforcement learning algorithm. In our model, the system evolves to minimize both the collision rate and the access delay. The experimental results show that our scheme improves system performance in terms of the access success rate, the failure rate, the collision rate, and the access delay.

1. Introduction

In smart city applications, many smart and mobile devices connect with one another and operate in an adaptive manner. The devices generate relevant city data in bursts and in unexpected manners on a massive scale. Because Long Term Evolution-Advanced (LTE-A) cellular networks provide wide coverage and low latency, LTE-A is considered to be one of the most promising communication infrastructures for smart city applications. However, radio resources in this communication infrastructure are too limited to serve a large amount of data. 3rd Generation Partnership Project (3GPP) specifies that Machine Type Communication (MTC) should be used to handle the congestion problem caused by small amounts of data being transmitted from a large number of devices within a short period of time [1–3]. MTC traffic includes periodic-update traffic and event-driven traffic. Event-driven triggering brings about bursts and unpredictable traffic flows, and these add to the congestion problem [4].

When a device has data to send, it needs to synchronize with a base station, that is, Evolved Node B (eNB), and to reserve a Random Access Channel (RACH). The RACH is a sequence of physical radio resources (RA slots). To do this, a device follows a four-step random access procedure [5, 6]. First, a device with data to send selects a random access preamble randomly as a digital signature from a predefined set of preambles and sends the preamble to eNB. Then, eNB responds with a Random Access Response (RAR) message to synchronize subsequent uplink transmission. However, if multiple devices send the same preamble in the same RA slot at the first step, a collision occurs, and they will receive the same RAR message. After receiving the RAR message, a device sends a connection request message along with a scheduling request. Finally, eNB acknowledges the connection request message. If a device receives the acknowledgement message successfully, it proceeds to data transmission. If a device encounters a preamble collision, it does not receive the message from eNB and will initiate a new RA procedure after a fixed backoff time. When the number of unsuccessful attempts of a device reaches the predefined maximum value (*Maximum Number of Preamble Transmission* [1]), the device finally fails the RA process.

As the number of devices attempting random access in the same RA slot increases, the numbers of preamble collisions and access delays increase as well. The access delay is the time between the generation of access request and the completion of the random access procedure. As a result of such a delay, the congestion becomes heavy. 3GPP specifies the use of the Access Class Barring (ACB) scheme to tackle the congestion problem [7]. ACB is a well-known scheme that restricts RA attempts. ACB operates on two strategic parameters: the barring factor and the barring duration. Based on the current congestion status, eNB regulates the RA attempts of MTC devices using these two parameters. Thus, the control of the parameters is vital in protecting the system from the excessive connectivity of a large number of devices. However, 3GPP does not specify how to control the parameters dynamically.

In this paper, we model the problem of determining ACB parameter values by using a reinforcement learning algorithm. This algorithm is able to follow unexpected changes and traffic bursts rapidly. Through the learning algorithm, we propose a scheme for dynamically and autonomously controlling the parameters. The experimental results show that the use of our scheme is sufficient to resolve the congestion problem in terms of access success rate, failure rate, collision rate, and access delay.

The rest of the paper is organized as follows. In Section 2, we discuss related studies. In Section 3, we propose an access management scheme that uses a reinforcement learning algorithm. In Sections 4 and 5, we evaluate the performance of our scheme and conclude the paper with our plans for future research.

2. Related Work

Several proposals for tackling the RA congestion problem are discussed. In ACB [7], eNB broadcasts the barring factor ($0 \leq p \leq 1$) and barring duration to its cell based on the current congestion level. A device with data to send generates a random number ($0 \leq q \leq 1$). If $q \leq p$, the device gets permission to access RACH. Otherwise, the RA attempt is barred for the barring duration. However, there is tradeoff with respect to barring factor p. If severe congestion occurs in a cell, eNB sets p to an extremely low value and most devices are barred. This results in an unacceptable access delay. On the other hand, if eNB sets p to an extremely high value, most of the preambles encounter collisions. This results in unacceptable data transmission failure. Thus, the barring factor is an important factor in determining system performance.

3GPP specifies the use of Extended ACB (EAB) as well as ACB. In EAB [8], devices are grouped into a set of ACs (Access Classes). eNB broadcasts a barring bitmap for the ACs periodically. A device with data to send compares its AC with the bitmap. If the bit that corresponds to the AC of the device is set, the device is barred from transmitting data until the bit changes. In this case, the scheduling policy among the ACs is a factor determining system performance. In current cellular networks, eNB alone determines the ACB barring factor to stabilize each cell. In [9], a cooperative mechanism is proposed to control congestion globally over multiple cells. The barring factor of each eNB is decided cooperatively among all eNBs. This is done for global stabilization and for access load sharing.

In order to maintain a high service quality for HTC (Human Type Communication), 3GPP specifies the use of two different schemes: the MTC specific backoff scheme and the separate RA resources scheme [10]. In the MTC specific backoff scheme, a dedicated backoff parameter is set for the MTC devices. The backoff scheme discourages the devices to attempt random access for certain duration of time. The backoff value for HTC devices is shorter than it is for MTC devices. In the separate RA resources scheme, RA slots are allocated to HTC and MTC devices separately. Both of the schemes focus on reducing the impact of RACH congestion on HTC devices. Thus, MTC devices may experience serious congestion because the amount of resources available is reduced.

In addition to the solutions specified by 3GPP, various other congestion solutions are proposed. In [11], a congestion-aware admission control scheme is proposed. It rejects RA requests from MTC devices selectively according to the congestion level, which is directly induced from the incoming packet processing delay at the application layer. In [12], RACH resources are preallocated to different MTC classes using class-dependent backoff procedures to prevent a large number of simultaneous RACH access attempts. Dynamic access barring according to the traffic load level is proposed for collision avoidance. Under this barring, the access attempts of devices transmitting for the first time are delayed.

3. Proposed Scheme

To tackle the congestion problem, we adopt Q-learning (QL) algorithm. The algorithm utilizes a form of reinforcement learning to solve Markovian decision problems without possessing complete information [13, 14]. Because QL finds solutions through the experience of interacting with an environment, we use it to model the ACB barring factor [15]. In other words, we control the ACB barring factor adaptively with QL.

Let S denote a finite set of possible environment states and let A denote a finite set of admissible actions to be taken. At RA slot t, eNB perceives the current state $s_t = s \in S$ of the environment and takes an action $a_t = a \in A$ based on both the perceived state and its past experience. The action a_t changes the environmental state from s_t to $s_{t+1} = s' \in S$. When that happens, the system receives the reward r_t.

The goal of the QL algorithm is to find an optimal policy for state s that optimizes the rewards over the long run. The algorithm estimates the Q-value $Q(s, a)$ as the cumulative discounted reward. Using the Q-values, the algorithm finds the optimal Q-value $Q^*(s, a)$ in a greedy manner. The Q-value is updated as

$$Q(s, a) = Q(s, a) + \alpha \cdot \Delta Q(s, a), \tag{1}$$

where α ($0 \leq \alpha \leq 1$) is the learning rate. When α is 0, the Q-value is not updated. When α is a high value, learning occurs quickly, as in

$$\Delta Q(s, a) = \left\{ r + \gamma \cdot \max_{a' \in A} Q\left(s', a'\right) \right\} - Q(s, a), \tag{2}$$

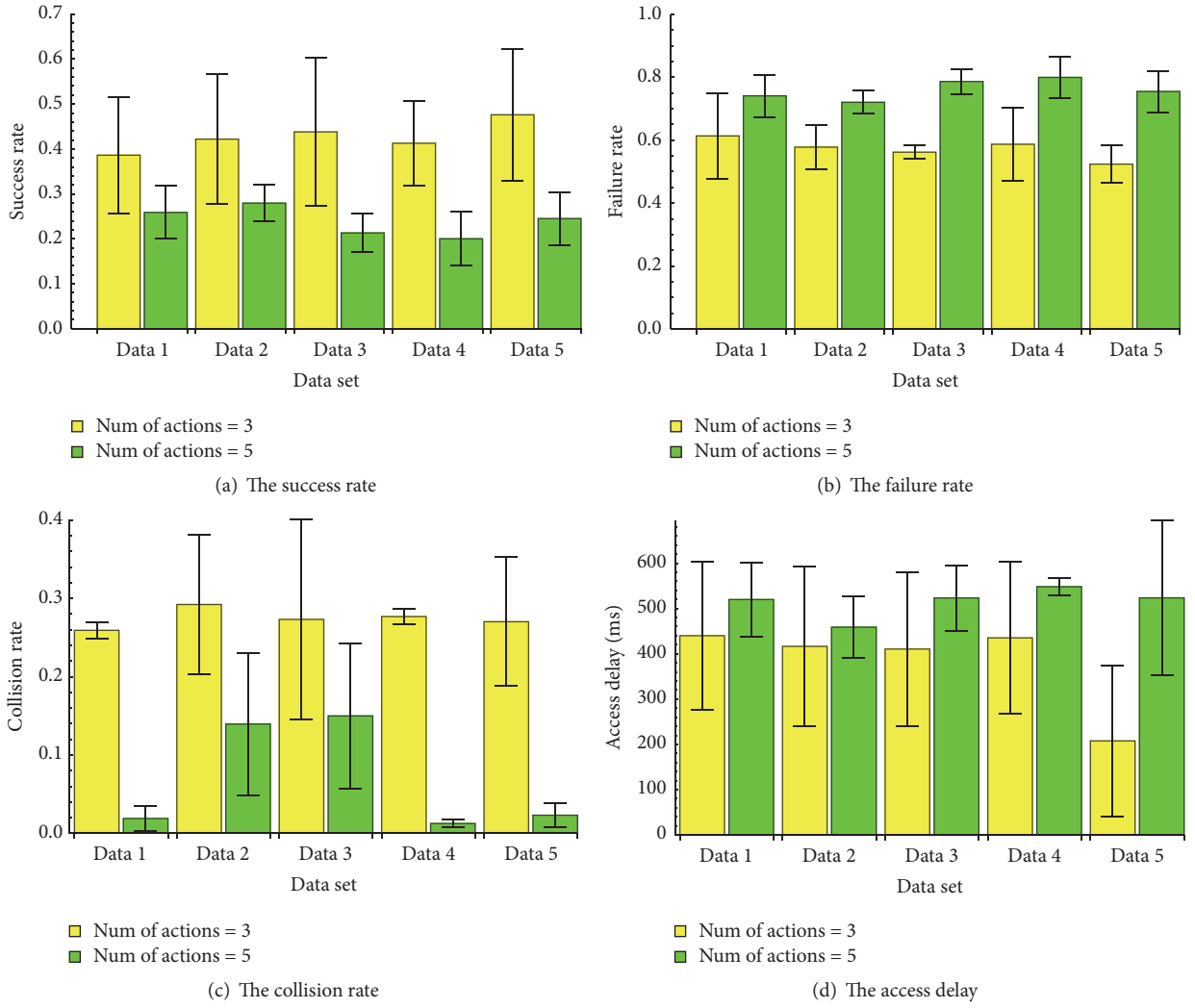


Figure 1: A performance comparison for various number of admissible actions.

where γ ($0 \leq \gamma \leq 1$) is the discount factor that weighs immediate rewards more heavily than future rewards.

In this paper, we model a QL algorithm to control the barring factor in order to minimize both the number of RACH collisions and the access delay. A collision occurs when two or more devices transmit the same preamble in the same time slot. We define a set of possible states, a set of admissible actions, and the rewards. First, we define the access success rate R_t^{succ} ($0 \leq R_t^{\text{succ}} \leq 1$) as a set with states S. The access success rate is defined as the number of devices that successfully access RACH divided by the number of devices contending in a given RA slot. R_t^{succ} is divided evenly into $|S|$ states. Each state s has three possible actions: increasing or decreasing the p value by $\delta_i \in \Delta$ or maintaining the current p value. The Δ indicates a finite set of unit values for p. To balance the exploration and exploitation of learning, an ϵ-greedy method [16] is applied to our QL algorithm. In other words, we can select a random action with probability ϵ or we can select an action with probability $1 - \epsilon$ that gives an optimal $Q(s, a)$ in the state s. We define the reward in order to minimize the collision rate (R_t^{col}, $0 < R_t^{\text{col}} < 1$) and the access delay (delay_t, $0 < \text{delay}_t < \text{delay}_{\max}$). R_t^{col} represents the number of colliding devices divided by the number of contending devices. The reward given when action a is taken at state s in RA slot t is

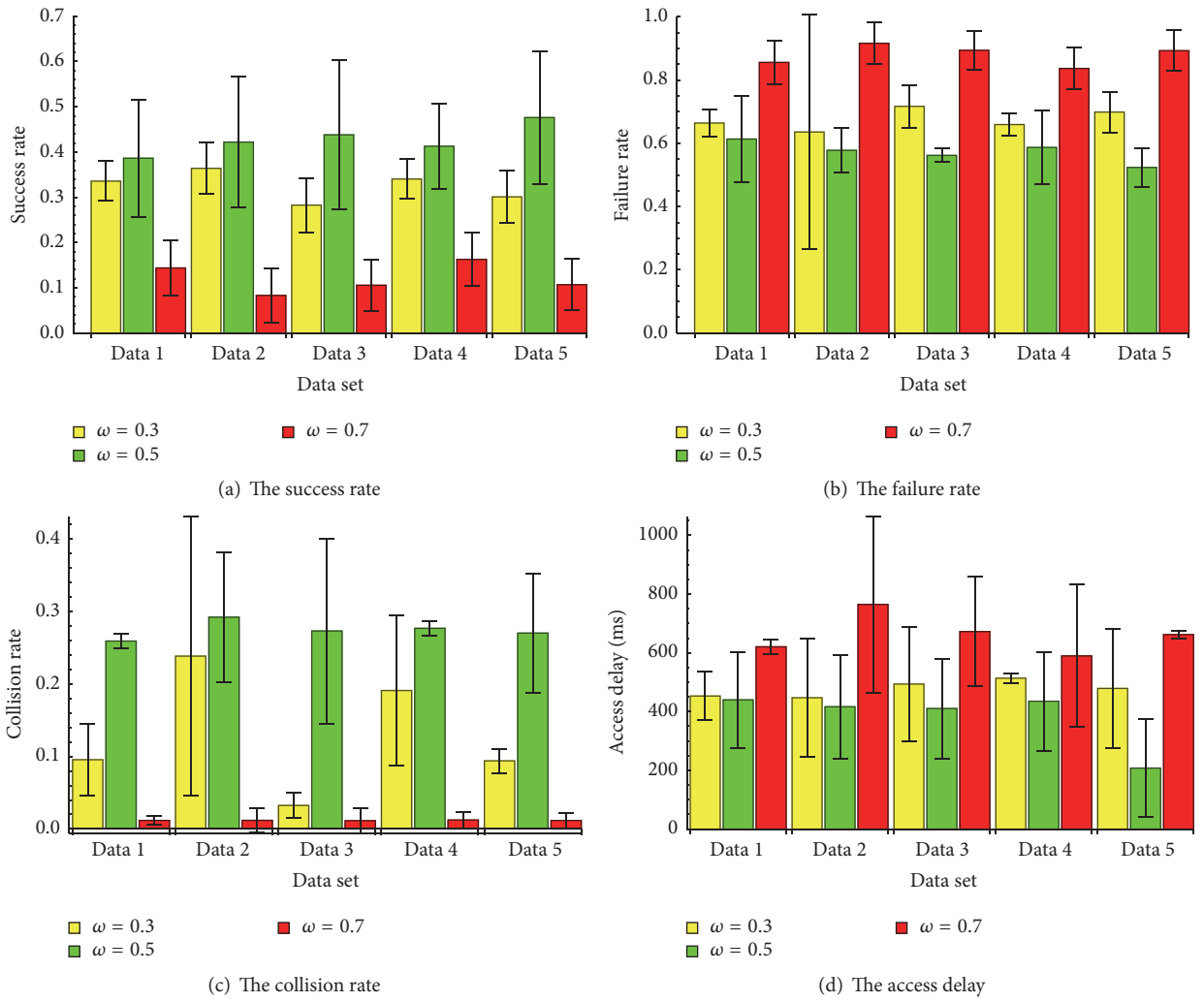
$$r_t(s, a) = \omega \cdot \frac{1}{R_t^{\text{col}}} + (1 - \omega) \frac{\text{delay}_{\max}}{\text{delay}_t}, \tag{3}$$

where $\text{delay}_{\max}$ is the maximum access delay that the system allows and ω is a smoothing factor ($0 \leq \omega \leq 1$).

4. Performance Analysis

In this paper, we extend the work of our previous study [15] and evaluate the performance of our access management scheme in terms of access success rate, collision rate, failure rate, and access delay.

We adopted the traffic model for a smart metering application as an experimental scenario in which a large



(a) The success rate

(b) The failure rate

(c) The collision rate

(d) The access delay

Figure 2: A performance comparison for various values of ω.

number of devices access RACHs in a highly synchronized manner [1]. Smart metering is one of smart city applications. In the model, the housing density of an urban area of London located within a single cell was used as the density of meters. We set the number of meters N to 35,670. Each meter requested one data transmission for reading frequency T. We set the reading frequency to 5 min. 3GPP defines two different traffic models for smart metering applications. We adopted a Beta distribution based model for our experiments. The number of meters that start the RA procedure in the tth RA slot is defined as

$$n_t = N \int_t^{t+1} p(t)\, dt, \tag{4}$$

where $p(t)$ follows the Beta distribution. The $p(t)$ is defined as

$$p(t) = \frac{t^{\alpha} \cdot (T-t)^{(\beta-1)}}{T^{\alpha+\beta-1} \cdot \mathrm{Beta}(\alpha, \beta)}, \tag{5}$$

where $\mathrm{Beta}(\alpha, \beta)$ is the Beta function with $\alpha = 3$ and $\beta = 4$.

In our QL model, we divided R_t^{succ} evenly into 4 states: s_1 for $0 \le R_t^{\mathrm{succ}} < 0.25$, s_2 for $0.25 \le R_t^{\mathrm{succ}} < 0.5$, s_3 for $0.5 \le R_t^{\mathrm{succ}} < 0.75$, and s_4 for $0.75 \le R_t^{\mathrm{succ}} \le 1$. $\mathrm{delay}_{\max}$ was set to the size of an RA slot multiplied by the *Maximum Number of Preamble Transmissions*. We set the learning rate α in (1) to 0.9. For the ϵ-greedy method, we set ϵ to 0.01. The basic RACH capacity parameters for LTE FDD networks followed [1].

Our scheme was trained using five different datasets. Each dataset included four training sets and one test set. The training sets contained the RA requests generated by the meters in a series of four reading frequencies. After the training, we measured the performance metrics of the test set. In the figures, the plotted values indicate the averages of values measured from the five datasets.

Figure 1 shows the scheme's performance with respect to various numbers of admissible actions. We defined the set elements $\Delta = \{\delta_1 = 0.2, \delta_2 = 0.1\}$ as operators to be used in actions. We used only δ_1 when there were three admissible actions: increasing p by δ_1, decreasing p by δ_1, and maintaining the current value of p. We used both δ_1

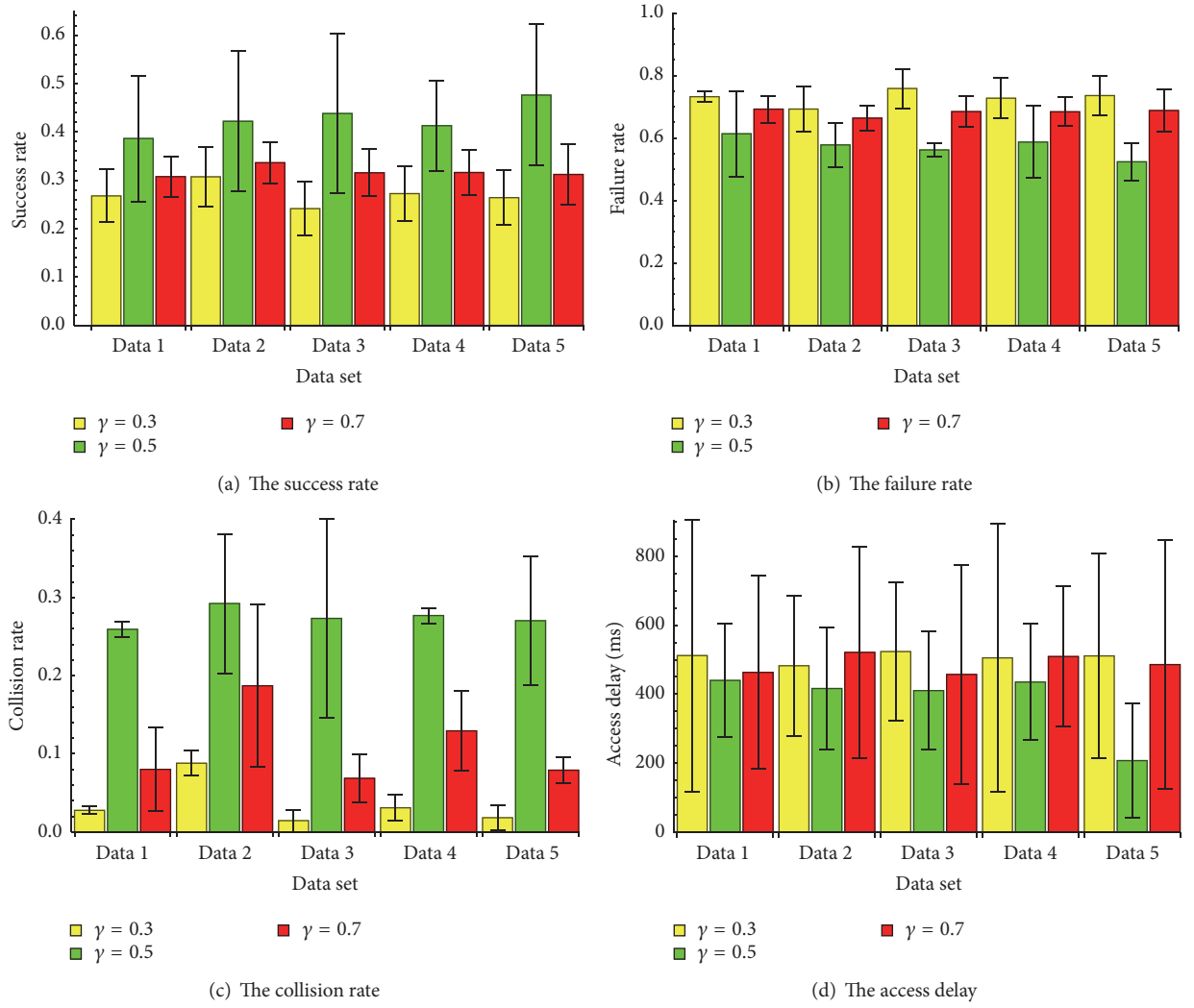


FIGURE 3: A performance comparison for various values of γ.

and δ_2 when there were five admissible actions: increasing p by δ_1 or δ_2, decreasing p by δ_1 or δ_2, and maintaining the current value of p. For the reward in (3), we set ω to 0.5 and we set the discount factor γ in (2) to 0.5. For the access success rate, the failure rate, and the access delay, the scheme with three admissible actions showed about 78%, 25%, and 26% better performances, respectively, than the scheme with five admissible actions did. The failure rate was calculated by taking the number of devices that ultimately failed RA attempts because the preamble transmission counter had reached *Maximum Number of Preamble Transmission* and dividing it by the number of contending devices. The scheme with three admissible actions showed a collision rate of about 4 times that of the scheme with five admissible actions. The performance with respect to various numbers of admissible actions is mainly influenced by the granularity of δ. When the granularity is properly coarse (e.g., $\delta = 0.2$), the barring factor swiftly copes with the variance of the number of meters trying to access RACH. However, when the granularity is too fine (e.g., $\delta = 0.1$), the barring factor does not promptly respond to the variance of RA requests. The issue to determine the level of granularity is still open.

Figure 2 shows the scheme's performance with respect to various values of ω, which influenced the rewards, as shown in (3). We considered three admissible actions involving δ_1 with $\gamma = 0.5$. For the access success and failure rates, the scheme with $\omega = 0.5$ showed about 30% and 16% better performances than those of the scheme with $\omega = 0.3$. Moreover, the scheme with $\omega = 0.5$ showed about 3 times and 35% better performance than the scheme with $\omega = 0.7$ did. For collision rate, the scheme with $\omega = 0.7$ showed about 11 times and 24 times better performances than those with $\omega = 0.3$ and $\omega = 0.5$, respectively. This is because the rewards added more weight to the collision rate when $\omega = 0.7$. For access delay, the scheme with $\omega = 0.5$ showed about 25% and 73% better performances than those of the schemes with $\omega = 0.3$ and $\omega = 0.7$, respectively. In these cases, the rewards added weight to access delay, and the scheme with $\omega = 0.3$ showed about 38% better performance than that of the scheme with $\omega = 0.7$. As shown in the figure, the performance

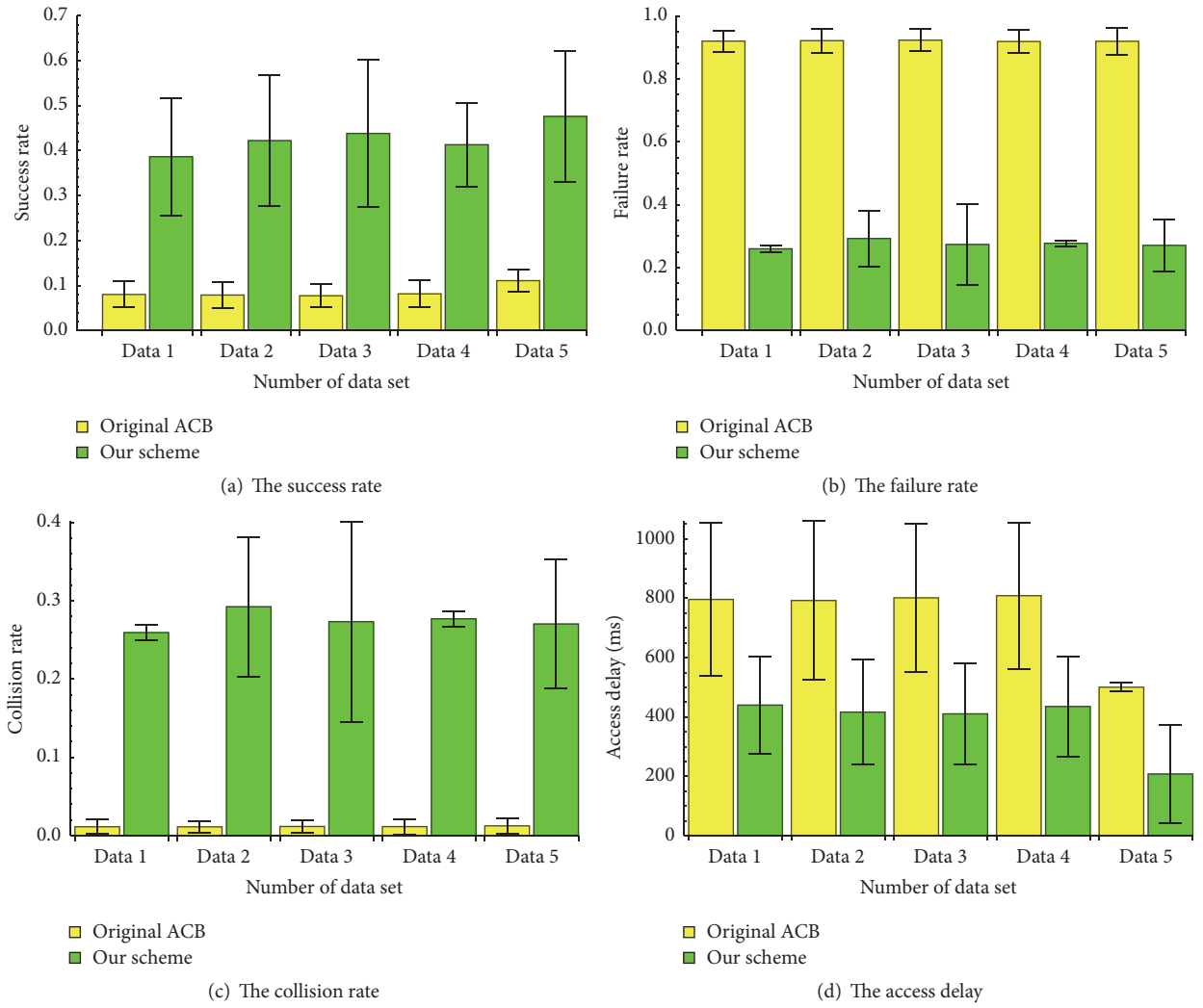


FIGURE 4: A performance comparison with the original ACB.

is significantly influenced by ω. When the weight is more added to the collision rate, the collision rate is improved. When the weight is more added to access delay, the access delay is improved. The mechanism to dynamically control the smoothing factor needs to improve the performance and we leave it for our future research.

Figure 3 shows the scheme's performance with respect to various values for the discount factor γ in (2). We used three admissible actions involving δ_1 with $\omega = 0.5$. When γ was low, weight was added to quicker rewards. For the success rate, failure rate, and access delay, the scheme with $\gamma = 0.5$ showed about 37%, 22%, and 25% better performances than the scheme with $\gamma = 0.3$ did. Compared to the scheme with $\gamma = 0.7$, the scheme with $\gamma = 0.5$ showed about 26%, 17%, and 22% better performances, respectively. For collision rate, the scheme with $\gamma = 0.3$ showed about 88% and 68% better performances than those of the schemes with $\gamma = 0.5$ and $\gamma = 0.7$, respectively.

To evaluate the scheme's performance, we compared our scheme to the original ACB [7]. In the original ACB, when congestion is detected, eNB regulates meters' RA attempts by setting the barring factor to 0.1. For our scheme, we used three admissible actions involving δ_1 with $\omega = 0.5$ and $\gamma = 0.5$. In Figure 4, our scheme shows a success rate of about 5 times better than that of the original ACB. In terms of the failure rate and access delay, our scheme showed about 70% and 50% better performances, respectively. For collision rate, the original ACB showed a very low value compared to that of our scheme. In the ACB, because most meters have restricted access to RACH and the competition for channel resources is reduced, the success and collision rates decrease, but both the failure rate and the access delay increase.

5. Conclusion

To tackle the RA congestion problem of MTC in LTE-A networks, we modelled the ACB barring factor decision problem using a Q-learning algorithm. The algorithm was able to follow unexpected and burst traffic changes rapidly. The goals of our model were minimizing both the RACH

collision rate and the access delay. To meet these goals, we defined sets of possible states and of admissible actions by using both the access success rate and the unit values to change the barring factor. To evaluate the performance of our scheme, we adopted the traffic model for smart metering applications. The results show that our scheme improves system performance in terms of the access success rate, the failure rate, the collision rate, and the access delay.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the Sookmyung Women's University Research Grants (1-1703-2030) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A09057141).

References

[1] 3rd Generation Partnership Project, "Technical specification group radio access network; study on RAN improvements for machine-type communications," 3GPP TR 37.868 V11.0.0, 2011.

[2] 3rd Generation Partnership Project, "Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," 3GPP TS 36.331 V12.7.0, 2015.

[3] M. S. Ali, E. Hossain, and D. I. Kim, "LTE/LTE-A random access for massive machine-type communications in smart cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 76–83, 2017.

[4] P. Jain, P. Hedman, and H. Zisimopoulos, "Machine type communications in 3GPP systems," *IEEE Communications Magazine*, vol. 50, no. 11, pp. 28–35, 2012.

[5] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A. C.-C. Hsu, "Overload control for machine-type-communications in LTE-advanced system," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 38–45, 2012.

[6] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 4–16, 2014.

[7] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 86–93, 2013.

[8] Intel Corporation, "Further performance evaluation of EAB information update mechanisms," 3GPP R2-120270, RANWG2 Meeting #77, 2012.

[9] S.-Y. Lien, T.-H. Liau, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 27–32, 2012.

[10] 3rd Generation Partnership Project, "RACH overload solutions," 3GPP R2-103742, RANWG2 #70bis, 2010.

[11] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: overload control," *IEEE Network*, vol. 26, no. 6, pp. 54–60, 2012.

[12] J.-P. Cheng, C.-H. Lee, and T.-M. Lin, "Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks," in *Proceedings of the IEEE GLOBECOM Workshops (GC Wkshps '11)*, pp. 368–372, Houston, Tex, USA, December 2011.

[13] R. S. Sutton, A. G. Barto, and R. J. Williams, "Reinforcement Learning is Direct Adaptive Optimal Control," *IEEE Control Systems*, vol. 12, no. 2, pp. 19–22, 1992.

[14] A. G. Barto, S. J. Bradtke, and S. P. Singh, "Learning to act using real-time dynamic programming," *Artificial Intelligence*, vol. 72, no. 1-2, pp. 81–138, 1995.

[15] J. Moon and Y. Lim, "Access control of MTC devices using reinforcement learning approach," in *Proceedings of the IEEE International Conference on Information Networking (ICOIN '17)*, pp. 641–643, 2017.

[16] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 3194845, 11 pages
https://doi.org/10.1155/2017/3194845

Research Article

Lightweight Data Aggregation Scheme against Internal Attackers in Smart Grid Using Elliptic Curve Cryptography

Debiao He,[1,2] Sherali Zeadally,[3] Huaqun Wang,[4] and Qin Liu[1]

[1]*State Key Lab of Software Engineering, Computer School, Wuhan University, Wuhan, China*
[2]*Co-Innovation Center for Information Supply & Assurance Technology, Anhui University, Hefei, China*
[3]*College of Communication and Information, University of Kentucky, Lexington, KY, USA*
[4]*Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, China*

Correspondence should be addressed to Qin Liu; qinliu@whu.edu.cn

Received 17 February 2017; Accepted 30 March 2017; Published 9 May 2017

Academic Editor: Jaime Lloret

Recent advances of Internet and microelectronics technologies have led to the concept of smart grid which has been a widespread concern for industry, governments, and academia. The openness of communications in the smart grid environment makes the system vulnerable to different types of attacks. The implementation of secure communication and the protection of consumers' privacy have become challenging issues. The data aggregation scheme is an important technique for preserving consumers' privacy because it can stop the leakage of a specific consumer's data. To satisfy the security requirements of practical applications, a lot of data aggregation schemes were presented over the last several years. However, most of them suffer from security weaknesses or have poor performances. To reduce computation cost and achieve better security, we construct a lightweight data aggregation scheme against internal attackers in the smart grid environment using Elliptic Curve Cryptography (ECC). Security analysis of our proposed approach shows that it is provably secure and can provide confidentiality, authentication, and integrity. Performance analysis of the proposed scheme demonstrates that both computation and communication costs of the proposed scheme are much lower than the three previous schemes. As a result of these aforementioned benefits, the proposed lightweight data aggregation scheme is more practical for deployment in the smart grid environment.

1. Introduction

By providing bidirectional communications of electricity and information, the smart grid performs real-time monitoring of power usage [1]. Based on the real-time information, the providers can monitor the power generation and consumption and get immediate power demand of each area. Then, they can take prompt action to optimize the power supply. The consumer can also get the current power price and adjust his/her behavior to lower expenses. Therefore, the smart grid can achieve efficient, economical, and reliable power services. Due to such advantages, the smart grid was a widespread concern for governments, industry, and academia in the last decade and is considered as the most promising candidate of the next generation power system [2].

The National Institute of Standards and Technology (NIST) presents a model and describes seven important domains of the smart gird [3]. As shown in Figure 1 [4], a smart gird consists of seven important domains, that is, the power generation (PG) domain, the power transmission (PT) domain, the power distribution (PD) domain, the power customer (PC) domain, the power operation (PO) domain, the power market (PM) domain, and the power service provider (PSP) domain [5, 6]. After being generated, transmitted, and distributed in the PG domain, the PT domain, and the PD domain, respectively, the customers in the PC domain can enjoy wonderful life based on the power. The PO domain, the PM domain, and the PSP domain manage the power flow, the participants, and all third-party operations, respectively [7, 8].

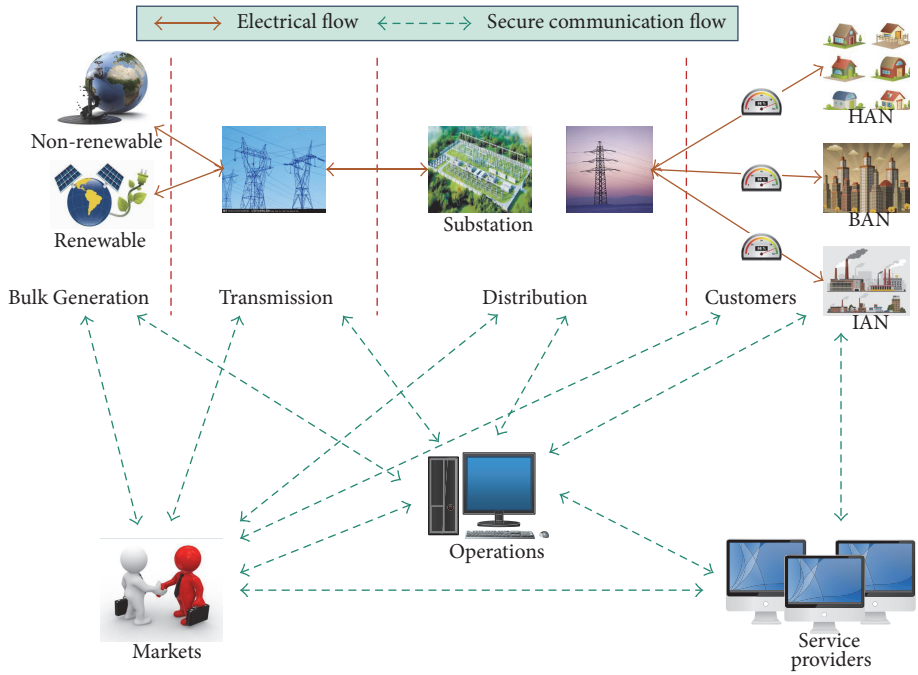


Figure 1: The model of the smart grid.

The smart meters in the smart grid collect the consumers' power consumption data and other information and send them to the remote control center. Generally speaking, the smart meter is installed outside the door of a consumer and an attacker is in charge of the communication channel easily due to its openness. The attacker may maliciously modify the power consumption data to increase/decrease the consumer's power expense. He/she also can get the daily routine of the consumer in order to commit crimes. For example, he/she knows that the consumer goes out when there is no power consumption and sneaks into the house to steal expensive things.

To address the above problems, how to achieve secure communications in the smart grid becomes an issue that needs to be addressed. In particular, ensuring the data's integrity and confidentiality is even more important. Several cryptographic schemes can be applied for secure communications in the smart grid. Many key management schemes [9–11], key distribution schemes [12–14], and key agreement schemes [15–17] were presented in recent years. However, many of these schemes cannot implement the integrity and confidentiality simultaneously. To address this challenge, data aggregation schemes have been proposed by several researchers and applied in the smart grid. However, most of them are vulnerable to attacks from internal attackers. Although several data aggregation schemes against internal attackers were proposed to enhance security, their computation or communication costs are too high for practical smart grid applications. In addition, the smart meter has very limited computation and communication capabilities. It is therefore necessary to design lightweight data aggregation schemes for practical deployment.

1.1. Our Contributions. To reduce both computation and communication costs, we propose a lightweight data aggregation scheme based on the Elliptic Curve Cryptography (ECC) [18, 19], which can obtain the same security level but with a much shorter key size. The main contributions of our paper are demonstrated as follows:

(i) First, we propose a lightweight data aggregation scheme based on Schnorr's signature scheme [18].

(ii) Second, we prove that the proposed lightweight data aggregation scheme is secure and is able to satisfy security requirements.

(iii) Finally, we analyze the performance of the proposed lightweight data aggregation scheme to demonstrate its high performance.

1.2. Organization of the Paper. In Section 2, we briefly review related papers about data aggregation schemes. In Section 3, we give some preliminaries, including backgrounds of ECC, network model, and security requirements of the data aggregation scheme. In Section 4, we present our lightweight data aggregation scheme based on ECC. In Section 5, we describe a security model for the data aggregation scheme and present the security analyses of our scheme. In Section 6, we present the computation and communication analyses of our data aggregation scheme.

2. Related Works

To guarantee secure communication in open environments, a lot of authentication schemes [20–22], encryption schemes [23–26], and secure outsourcing schemes [25, 27, 28] have

been constructed in last several years. Li et al. [29] and Garcia and Jacobs [30] designed two data aggregation schemes using Paillier's encryption scheme [31]. To improve performance, Lu et al. [32] designed an improved data aggregation scheme using Paillier's encryption scheme and the super-increasing sequence. However, the above three schemes [29, 30, 32] cannot protect consumers' privacy because none of them can provide anonymity. To protect consumers' privacy, Zhang et al. [33] designed a security-enhanced data aggregation scheme based on the Chinese Remainder Theorem and Paillier's encryption scheme. Chen et al. [34] also designed a security-enhanced data aggregation scheme with fault tolerance based on Paillier's encryption scheme.

Unfortunately, internal attacks are not considered in the above data aggregation schemes [29, 30, 32–34] thereby allowing internal attackers to access the consumers smart grid data. To address this weakness, Fan et al. [35] designed the first data aggregation scheme that can withstand attacks from internal attackers by using blinding technology. Unfortunately, Bao and Lu [36] demonstrated that Fan et al.'s data aggregation scheme cannot guarantee the integrity of transmitted data. To enhance security, He et al. [4] designed an improved data aggregation scheme based on Boneh et al.'s encryption scheme [37]. The performance of Fan et al.'s data aggregation scheme [35] and He et al.'s data aggregation scheme [4] is not good enough because they use bilinear pairing operations.

3. Preliminaries

3.1. Elliptic Curve. Given a prime number p, we say that the equation $y^2 = x^3 + a \cdot x + b \bmod p$ defines an elliptic curve $E(F_p)$, where $a, b \in F_p$ and $\Delta = 4a^3 + 27b^2 \neq 0 \bmod p$ [38]. It is well known that all points on $E(F_p)$ and the infinite point $\mathcal{O}$ make an additive group $\mathcal{G}$. Given a generator point P with a prime order q, the scale multiplication operation is defined as $n \cdot P = P + P + \cdots + P_{n \text{ times}}$, where n is a positive integer.

Previous researches have showed that the following problems in the group $\mathcal{G}$ are suitable for the design of public key cryptography because no probabilistic polynomial time algorithm can solve them efficiently [38].

Discrete Logarithm (DL) Problem. Given an element $Q \in \mathcal{G}$, the DL problem is to extract an element $x \in Z_q^*$ such that $Q = x \cdot P$.

Computational Diffie-Hellman (CDH) Problem. Given two elements $x \cdot P, y \cdot P \in \mathcal{G}$ with two unknown elements $x, y \in Z_q^*$, the CDH problem is to extract the element $Q = x \cdot y \cdot P$.

3.2. Network Model. As shown in Figure 2 [4], there are three participants in the system of a data aggregation scheme, namely, a trusted third party (TTP), an aggregator (Agg), and a smart meter (SM_i) [4, 35]. The functions of the above three participants are presented as below.

(i) TTP: it is a trusted third party and its function is to generate blinding factors to withstand the internal attackers.

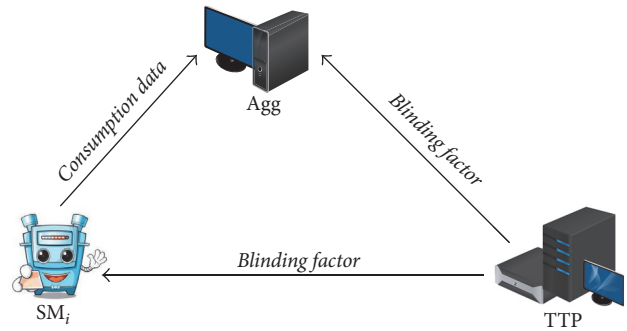


Figure 2: The registration phase.

(ii) Agg: it is the manager of the smart grid and its function is to generate the system parameters and the private keys of smart meters.

(iii) SM_i: it is a smart meter and its function is to collect consumers' electricity consumption data and send it to Agg.

The workflow of the system is presented as follows. (1) Agg produces the system parameters and the mast private key; (2) SM_i registers in Agg and gets its private key; (3) TTP generates the blinding factors for Agg and SM_i; (4) SM_i collects the electricity consumption, produces a ciphertext, and sends it to Agg; (5) after collecting all ciphertexts, Agg checks their validity and extracts the sum of all electricity consumption data.

3.3. Security Requirements. Based on recently works, we know that a data aggregation scheme for the smart grid should meet the below security requirements [4, 35].

(i) Confidentiality. The consumer's power consumption data indicates his/her habit and its leakage may be used by an attacker to commit a crime. To ensure the consumer's safety, a data aggregation scheme should provide confidentiality; that is, both the external attackers and the internal attackers cannot extract the electricity consumption data from intercepted messages.

(ii) Authentication. The malicious attacker may forge a message and impersonate the consumer. To ensure if the received message is transmitted by a legal SM_i, a data aggregation scheme should provide authentication; that is, Agg can check the legality of the received message.

(iii) Integrity. All messages are transmitted over open communication channels and the malicious attacker may modify them to break regular transactions. To protect the rights and interests of all participants in the smart grid, a data aggregation scheme should provide integrity; that is, Agg can detect any modification of the received data.

(iv) Resistance against Attacks. Due to the openness of communication channels in the smart grid, the system is vulnerable to many types of attacks. To obtain secure communications in the smart grid, a data aggregation scheme should

supply resistance against attacks; that is, it can withstand the replay attack, the modification attack, the man-in-the-middle attack, and the impersonation attack.

4. The Proposed Data Scheme

We describe our proposed lightweight data aggregation scheme, which consists of three phases, namely, the initialization phase, the registration phase, and the aggregation phase.

Initialization Phase. In this phase, Agg executes some steps to produce the system parameters. TTP and Agg execute some other steps to produce the blind factors against internal attackers.

Agg runs the following steps to produce the system parameters.

(1) Agg selects an elliptic curve $E(F_p)$ determined by the equation $y^2 = x^3 + a \cdot x + b \bmod p$, where p is a prime and $a, b \in Z_q$.

(2) Agg selects an element P with the order q existing on $E(F_p)$, where q is a prime.

(3) Agg selects an element $s \in Z_q^*$ and calculates $P_{\text{pub}} = s \cdot P$.

(4) Agg selects three cryptographic hash functions $h_i : \{0,1\}^* \to Z_q^*$ $(i = 1, 2, 3)$.

(5) Agg publishes params $= \{p, a, b, q, P, P_{\text{pub}}, h_1, h_2, h_3\}$ and saves s secretly.

TTP and Agg execute the following steps to produce the blinding factors.

(1) TTP randomly selects a group of elements $\theta_1, \theta_2, \ldots, \theta_n \in Z_q^*$ and computes $\theta = \sum_{i=1}^{n} \theta_i \bmod q$. At last, TTP sends θ to Agg and also sends θ_i to SM_i, where $i = 1, 2, \ldots, n$.

(2) Agg computes $\theta_0 = -\theta \bmod q$ and keeps it secretly.

Registration Phase. In this phase, SM_i registers in Agg. After registration, SM_i receives its private key and becomes a legal smart meter. As demonstrated in Table 1, SM_i and Agg run the following processes to finish the registration.

(1) SM_i randomly chooses an element $x_i' \in Z_{\widehat{q}}^*$, computes $X_i' = x_i' \cdot P$, and transmits $\{\text{id}_i, X_i'\}$ to Agg secretly.

(2) Agg randomly chooses an element $x_i'' \in Z_{\widehat{q}}^*$ and computes $X_i'' = x_i'' \cdot P$, $X_i = X_i' + X_i''$, $\alpha_i = h_1(\text{id}_i, X_i)$, and $s_i'' = s + \alpha_i \cdot x_i'' \bmod q$. At last, Agg sends $\{s_i'', X_i''\}$ to SM_i secretly.

(3) SM_i computes $X_i = X_i' + X_i''$, $\alpha_i = h_1(\text{id}_i, X_i)$, $s_i = s_i'' + x_i' \cdot \alpha_i \bmod q$ and checks if the equation $s_i \cdot P = P_{\text{pub}} + \alpha_i \cdot X_i$ holds. If not, SM_i rejects the session; otherwise, SM_i stores $\{s_i, X_i\}$ and finishes the registration.

Table 1: The registration phase of our scheme.

SM_i		Agg
Generate $x_i' \in Z_q^*$;		
$X_i' = x_i' \cdot P$		
	$\xrightarrow{\{\text{id}_i, X_i'\}}$	
		Generate $x_i'' \in Z_q^*$;
		$X_i'' = x_i'' \cdot P$;
		$X_i = X_i'' + X_i'$;
		$\alpha_i = h_1(\text{id}_i, X_i)$;
		$s_i'' = s + \alpha_i \cdot x_i'' \bmod q$
	$\xleftarrow{\{s_i'', X_i''\}}$	
$X_i = X_i'' + X_i'$;		
$\alpha_i = h_1(\text{id}_i, X_i)$;		
$s_i = s_i'' + \alpha_i \cdot x_i' \bmod q$;		
check $s_i \cdot P \stackrel{?}{=} P_{\text{pub}} + \alpha_i \cdot X_i$;		
store $\{s_i, X_i\}$		

Due to the fact that $X_i' = x_i' \cdot P$, $X_i'' = x_i'' \cdot P$, $X_i = X_i' + X_i''$, $s_i'' = s + \alpha_i \cdot x_i'' \bmod q$, and $s_i = s_i'' + x_i' \cdot \alpha_i \bmod q$, then we have

$$\begin{aligned} s_i \cdot P &= \left(s_i'' + x_i' \cdot \alpha_i\right) \cdot P = \left(s + \alpha_i \cdot x_i'' + \alpha_i \cdot x_i'\right) \cdot P \\ &= \left(s + \alpha_i \cdot \left(x_i' + x_i''\right)\right) \cdot P \\ &= s \cdot P + \alpha_i \cdot \left(x_i' + x_i''\right) \cdot P \\ &= P_{\text{pub}} + \alpha_i \cdot \left(x_i' \cdot P + x'' \cdot P\right) \\ &= P_{\text{pub}} + \alpha_i \cdot \left(X_i' + X_i''\right) = P_{\text{pub}} + \alpha_i \cdot X_i. \end{aligned} \tag{1}$$

Therefore, the correctness of the registration phase is demonstrated.

Aggregation Phase. In this phase, SM_i extracts the power consumption data and sends it to Agg. Agg checks the validity of the received messages and aggregates all the received data. As demonstrated in Table 1, the steps below are executed by SM_i and Agg.

(1) SM_i gets the power consumption data m_i, randomly chooses an element $y_i \in Z_{\widehat{q}}^*$, and computes $Y_i = y_i \cdot P$, $\widehat{Y}_i = y_i \cdot P_{\text{pub}}$, $c_i = m_i + \theta_i + h_2(\widehat{R}_i) \bmod q$, $\beta_i = h_3(\text{id}_i, X_i, Y_i, c_i, t)$, and $d_i = s_i + \beta_i \cdot y_i \bmod q$. At last, SM_i transmits $\{X_i, Y_i, c_i, d_i, t\}$ to Agg.

(2) Agg checks if $d_i \cdot P = P_{\text{pub}} + \alpha_i \cdot X_i + \beta_i \cdot Y_i$, where $\alpha_i = h_1(\text{id}_i, X_i)$ and $\beta_i = h_3(\text{id}_i, X_i, Y_i, c_i, t)$. To improve performance, we use the small exponent test technology [39] to achieve the batch verification. Agg randomly chooses a group of integers $z_1, z_2, \ldots, z_n \in [1, 2^w]$ and checks if the equation $(\sum_{i=1}^{n} z_i \cdot d_i) \cdot P = (\sum_{i=1}^{n} z_i) P_{\text{pub}} + \sum_{i=1}^{n} (z_i \cdot \alpha_i \cdot X_i + z_i \cdot \beta_i \cdot Y_i)$ holds. Agg computes $c = \sum_{i=1}^{n} (c_i - h_2(s \cdot Y_i))$ and extracts the sum of the power consumption data by computing $m = c + \theta_0 \bmod q$.

Due to the fact that $s_i \cdot P = P_{\text{pub}} + \alpha_i \cdot X_i$, $Y_i = y_i \cdot P$, $\widehat{Y}_i = y_i \cdot P_{\text{pub}}$, $c_i = m_i + \theta_i + h_2(\widehat{Y}_i) \bmod q$ and $d_i = s_i + \beta_i \cdot y_i \bmod q$, we can derive

$$\begin{aligned}
d_i \cdot P &= (s_i + \beta_i \cdot y_i) \cdot P = s_i \cdot P + \beta_i \cdot y_i \cdot P \\
&= P_{\text{pub}} + \alpha_i \cdot X_i + \beta_i \cdot Y_i \\
\left(\sum_{i=1}^{n} z_i \cdot d_i\right) \cdot P &= \left(\sum_{i=1}^{n} z_i \cdot (s_i + \beta_i \cdot y_i)\right) \cdot P \\
&= \left(\sum_{i=1}^{n} z_i \cdot (s_i + \beta_i \cdot y_i)\right) \cdot P \\
&= \sum_{i=1}^{n} (z_i \cdot s_i \cdot P + z_i \cdot \beta_i \cdot y_i \cdot P) \\
&= \sum_{i=1}^{n} \left(z_i \cdot \left(P_{\text{pub}} + \alpha_i \cdot X_i\right) + z_i \cdot \beta_i \cdot Y_i\right) \\
&= \left(\sum_{i=1}^{n} z_i\right) \cdot P_{\text{pub}} + \sum_{i=1}^{n} (z_i \cdot \alpha_i \cdot X_i + z_i \cdot \beta_i \cdot Y_i),
\end{aligned} \tag{2}$$

$$\begin{aligned}
c + \theta_0 &= \sum_{i=1}^{n} (c_i - h_2(s \cdot Y_i)) + \theta_0 \\
&= \sum_{i=1}^{n} \left(m_i + \theta_i + h_2\left(\widehat{R}_i\right) - h_2(s \cdot y_i \cdot P)\right) - \theta \\
&= \sum_{i=1}^{n} \left(m_i + \theta_i + h_2\left(\widehat{R}_i\right) - h_2(y_i \cdot s \cdot P)\right) - \theta \\
&= \sum_{i=1}^{n} \left(m_i + \theta_i + h_2\left(\widehat{R}_i\right) - h_2\left(y_i \cdot P_{\text{pub}}\right)\right) - \theta \\
&= \sum_{i=1}^{n} \left(m_i + \theta_i + h_2\left(\widehat{R}_i\right) - h_2\left(\widehat{R}_i\right)\right) - \theta \\
&= \sum_{i=1}^{n} m_i + \sum_{i=1}^{n} \theta_i - \sum_{i=1}^{n} \theta_i = \sum_{i=1}^{n} m_i.
\end{aligned} \tag{3}$$

According to the above equations, the correctness of the aggregation phase of our scheme is demonstrated.

5. Security Analysis

The security of the proposed lightweight data aggregation scheme is analyzed in this section. First, we present a security model for the data aggregation scheme. Second, we demonstrate that the proposed lightweight data aggregation scheme is provably secure in the security model. Finally, we demonstrate that the proposed lightweight data aggregation scheme can meet the security requirements presented in Section 3.

5.1. Security Model. Based on security models [40] for signcryption schemes, we presented a security model for data aggregation schemes. The security of confidentiality and unforgeability is formally defined by two games executed by an attacker $\mathcal{A}$ and a challenger $\mathcal{C}$. $\mathcal{A}$ is allowed to make the following queries.

(i) $h_i(m)$: for such a query made by $\mathcal{A}$, $\mathcal{C}$ randomly selects $r \in Z_q^*$, sends r to $\mathcal{A}$, and stores (m, r) in the table L_{h_i}, where $i = 1, 2, 3$.

(ii) CreateSM(id_i): for such a query made by $\mathcal{A}$, $\mathcal{C}$ generates SM_i's secret key and blinding factor and stores them in the table L_{SM}.

(iii) CorruptSM(id_i): for such a query made by $\mathcal{A}$, $\mathcal{C}$ sends SM_i's private key and blinding factor to $\mathcal{A}$.

(iv) Signcrypt(id_i, m_i): for such a query made by $\mathcal{A}$, $\mathcal{C}$ generates a ciphertext $\{X_i, Y_i, c_i, d_i, t\}$ corresponding to the message m_i.

(v) Designcrypt($\text{id}_i, X_i, Y_i, c_i, d_i, t$): for the query made by $\mathcal{A}$, $\mathcal{C}$ checks the validity of the ciphertext and decrypts it to get the plaintext.

Definition 1. A data aggregation scheme is able to provide confidentiality [indistinguishability against adaptive chosen ciphertext attacks (IND − CCA)] if no attacker can win the following game with a nonnegligible advantage.

Setup. $\mathcal{C}$ produces system parameters and transmits them to $\mathcal{A}$.

Phase 1. $\mathcal{A}$ is able to adaptively make h_i, CreateSM, CorruptSM, Signcrypt, and Designcrypt queries.

Challenge. $\mathcal{A}$ picks a challenging identity id_i^*, chooses two messages m_0 and m_1, and sends them to $\mathcal{C}$. $\mathcal{C}$ picks a random element $b \in \{0, 1\}$, produces a signcrypted ciphertext $\{X_i, Y_i, c_i, d_i, t\}$, and sends it to $\mathcal{A}$.

Phase 2. In this phase, $\mathcal{A}$ can adaptively make h_i, CreateSM, CorruptSM, and Signcrypt queries except that it cannot make a CorruptSM query with id_i^* or a Designcrypt query with $\{X_i, Y_i, c_i, d_i, t\}$.

Finally, $\mathcal{A}$ gives its guess $b' \in \{0, 1\}$ about the value of b selected by $\mathcal{C}$.

The advantage of $\mathcal{A}$ is defined by the equation $\text{Adv}_{\mathcal{A}}^{\text{IND-CCA}} = |2 \cdot \Pr[b' = b] - 1|$. $\mathcal{A}$ wins in the above game if it guesses the value of b correctly.

Definition 2. A data aggregation scheme is able to provide unforgeability [existential unforgeability against adaptive chosen messages attacks (EUFCMA)] if no attacker wins the following game with a nonnegligible advantage.

Setup. $\mathcal{C}$ produces the system parameters and sends them to $\mathcal{A}$.

Qurey. In this phase, $\mathcal{A}$ picks a challenging identity id_i^* and is able to adaptively make h_i, CreateSM, CorruptSM, Signcrypt, and Designcrypt queries except that it cannot make a CorruptSM query with id_i^*.

Forgery. In this phase, $\mathscr{A}$ outputs a ciphertext $\{X_i, Y_i, c_i, d_i, t\}$ corresponding to the challenging identity id_i^*.

We say $\mathscr{A}$ wins in the above game if $\{X_i, Y_i, c_i, d_i, t\}$ is valid and it is not generated by executing a Signcrypt query.

5.2. Security Analysis

Theorem 3. *The proposed data aggregation scheme is able to provide confidentiality if the CDH problem is hard.*

Proof. Assume that an attacker $\mathscr{A}$ wins the game defined in Definition 1 with a nonnegligible advantage ϵ. Based on $\mathscr{A}$'s capability, we can construct a challenger to solve the CDH problem with a nonnegligible advantage. Given an instance $(P, Q_1 = a \cdot P,\ Q_2 = b \cdot P)$ of the CDH problem, $\mathscr{C}$ sets $P_{\mathrm{pub}} \leftarrow a \cdot P$ and sends params $= \{p, a, b, q, P, P_{\mathrm{pub}}, h_1, h_2, h_3\}$ to $\mathscr{A}$. $\mathscr{C}$ randomly picks up an identity id_I as the challenging identity and answers queries from $\mathscr{A}$ according to the rules below.

(i) $h_i(m)$: $\mathscr{C}$ keeps a table L_{h_i} of the form (m, r), where $i \in \{1, 2, 3\}$. Upon receiving such a query, $\mathscr{C}$ checks if L_{h_i} contains a tuple (m, r). If so, $\mathscr{C}$ sends r to $\mathscr{A}$; otherwise, $\mathscr{C}$ randomly selects an element $r \in Z_q^*$, stores (m, r) into L_{h_i}, and sends r to $\mathscr{A}$.

(ii) CreateSM(id_i): $\mathscr{C}$ keeps a table L_{SM} of the form $(\mathrm{id}_i, \theta_i, s_i, X_i)$. Upon receiving such a query, $\mathscr{C}$ checks if L_{SM} contains a tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$. If so, $\mathscr{C}$ sends X_i to $\mathscr{A}$; otherwise, $\mathscr{C}$ randomly selects three integers $\theta_i, \alpha_i, s_i \in Z_q^*$ and sets $X_i \leftarrow \alpha_i^{-1} \cdot (s_i \cdot P - P_{\mathrm{pub}})$. $\mathscr{C}$ stores $(\mathrm{id}_i, X_i, \alpha_i)$ and $(\mathrm{id}_i, \theta_i, s_i, X_i)$ into L_{SM}, respectively.

(iii) CorruptSM(id_i): $\mathscr{C}$ checks if L_{SM} contains a tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$. If not, $\mathscr{C}$ makes CreateSM-query with the identity id_i. After that, $\mathscr{C}$ returns $(\mathrm{id}_i, \theta_i, s_i, X_i)$ to $\mathscr{C}$.

(iv) Signcrypt(id_i, m_i): $\mathscr{C}$ checks if L_{SM} contains a tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$. If not, $\mathscr{C}$ makes CreateSM-query with the identity id_i. After that, $\mathscr{C}$ gets the tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$ from L_{SM} and uses it to produce a ciphertext $\{X_i, Y_i, c_i, d_i, t\}$. At last, $\mathscr{C}$ sends $\{X_i, Y_i, c_i, d_i, t\}$ to $\mathscr{A}$.

Given the power consumption data m_0 and m_1, $\mathscr{C}$ extracts $(\mathrm{id}_I, \theta_I, s_I, X_I)$ from L_{SM} and selects a random element $b \in \{0, 1\}$. $\mathscr{C}$ sets $Y_I \leftarrow b \cdot P$, randomly selects three elements $\beta_I, c_I, d_I \in Z_q^*$, stores $(\mathrm{id}_I, X_I, Y_I, c_I, t, \beta_I)$ into L_{h_3}, and sends $\{X_I, Y_I, c_I, d_I, t\}$ to $\mathscr{A}$.

After that, $\mathscr{A}$ can make h_i, CreateSM, CorruptSM, and Signcrypt queries and get the corresponding responses. Then, $\mathscr{A}$ outputs b' as his/her guess against the confidentiality. $\mathscr{C}$ selects a random tuple (R, r) from L_{h_2} and outputs R as the solution of the given CDH problem.

Let q_{h_2} denote the number of h_2-query involved in the game. The probability that $\mathscr{C}$ can solve the given CDH problem is $\eta = \epsilon / q_{h_2}$. Because of the nonnegligibility of ϵ, we know that η is nonnegligible. This contradicts with the hardness of the CDH problem. Thus, the proposed data aggregation scheme is able to provide confidentiality. □

Theorem 4. *The proposed data aggregation scheme is able to provide unforgeability if the DL problem is hard.*

Proof. Assume that an attacker $\mathscr{A}$ wins the game defined in Definition 1 with a nonnegligible advantage ϵ. Based on $\mathscr{A}$'s capability, we can construct a challenger to solve the DL problem with a nonnegligible advantage. Given an instance $(P, Q_1 = a \cdot P)$ of the DL problem, $\mathscr{C}$ picks a random integer $s \in Z_q^*$, computes $P_{\mathrm{pub}} = s \cdot P$, and sends params $= \{p, a, b, q, P, P_{\mathrm{pub}}, h_1, h_2, h_3\}$ to $\mathscr{A}$. $\mathscr{C}$ randomly selects an identity id_I as the challenging identity and answers queries from $\mathscr{A}$ according to the rules below.

(i) $h_i(m)$: $\mathscr{C}$ keeps a table L_{h_i} of the form (m, r), where $i \in \{1, 2, 3\}$. Upon receiving such a query, $\mathscr{C}$ checks if L_{h_i} contains a tuple (m, r). If so, $\mathscr{C}$ sends r to $\mathscr{A}$; otherwise, $\mathscr{C}$ randomly picks up an element $r \in Z_q^*$, stores (m, r) into L_{h_i}, and sends r to $\mathscr{A}$.

(ii) CreateSM(id_i): $\mathscr{C}$ keeps a table L_{SM} of the form $(\mathrm{id}_i, \theta_i, s_i, X_i)$. Upon receiving such a query, $\mathscr{C}$ checks if L_{SM} contains a tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$. If so, $\mathscr{C}$ sends X_i to $\mathscr{A}$; otherwise, $\mathscr{C}$ answers the query through the rules below:

(1) If $\mathrm{id}_i = \mathrm{id}_I$, $\mathscr{C}$ randomly picks two integers $\theta_i, \alpha_i \in Z_q^*$ and sets $X_i \leftarrow a \cdot P$. $\mathscr{C}$ stores $(\mathrm{id}_i, X_i, \alpha_i)$ and $(\mathrm{id}_i, \theta_i, \perp, X_i)$ into L_{SM}, respectively.

(2) Otherwise ($\mathrm{id}_i \neq \mathrm{id}_I$), $\mathscr{C}$ randomly selects three integers $\theta_i, \alpha_i, s_i \in Z_q^*$ and sets $X_i \leftarrow \alpha_i^{-1} \cdot (s_i \cdot P - P_{\mathrm{pub}})$. $\mathscr{C}$ stores $(\mathrm{id}_i, X_i, \alpha_i)$ and $(\mathrm{id}_i, \theta_i, s_i, X_i)$ into L_{SM}, respectively.

(iii) CorruptSM(id_i): $\mathscr{C}$ checks if L_{SM} contains a tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$. If not, $\mathscr{C}$ makes CreateSM-query with the identity id_i. After that, $\mathscr{C}$ returns $(\mathrm{id}_i, \theta_i, s_i, X_i)$ to $\mathscr{A}$.

(iv) Signcrypt(id_i, m_i): $\mathscr{C}$ checks if id_i and id_I are equal. If they are not, $\mathscr{C}$ extracts the tuple $(\mathrm{id}_i, \theta_i, s_i, X_i)$ from L_{SM} and uses it to produce a ciphertext $\{X_i, Y_i, c_i, d_i, t\}$ according to the description of the proposed data aggregation; otherwise, $\mathscr{C}$ randomly selects two integers $d_i, \beta_i \in Z_q^*$ and computes $Y_i = \beta_i^{-1} \cdot (d_i \cdot P - P_{\mathrm{pub}} - \alpha_i \cdot X_i)$ and $c_i = m_i + \theta_i + h_2(s \cdot Y_i)$. $\mathscr{C}$ stores $(\mathrm{id}_i, X_i, Y_i, c_i, t)$ into L_{h_2} and sends $\{X_i, Y_i, c_i, d_i, t\}$ to $\mathscr{A}$.

(v) Designcrypt($\mathrm{id}_i, X_i, Y_i, c_i, d_i, t$): for the query made by $\mathscr{A}$, $\mathscr{C}$ checks the validity of the ciphertext and decrypts it to get the plaintext using the systems secret key s.

□

At last, $\mathscr{A}$ outputs a forged ciphertext $(\mathrm{id}_i, X_i, Y_i, c_i, d_i, t)$. $\mathscr{C}$ stops the game if the equation $\mathrm{id}_i = \mathrm{id}_I$ holds. Based on the forking lemma [41], $\mathscr{C}$ can output another valid ciphertext $(\mathrm{id}_i, X_i, Y_i, c_i, d_i', t)$ by choosing a different hash function h_1. Since both ciphertexts are valid, we can derive the following two equation:

$$\begin{aligned} d_i \cdot P &= P_{\mathrm{pub}} + \alpha_i \cdot X_i + \beta_i \cdot Y_i, \\ d_i' \cdot P &= P_{\mathrm{pub}} + \alpha_i' \cdot X_i + \beta_i \cdot Y_i. \end{aligned} \tag{4}$$

Based on the above two equations, we can derive the equation below:

$$\begin{aligned}\left(d_i - d_i'\right) \cdot P &= d_i \cdot P - d_i' \cdot P \\ &= \left(P_{\text{pub}} + \alpha_i \cdot X_i + \beta_i \cdot Y_i\right) \\ &\quad - \left(P_{\text{pub}} + \alpha_i' \cdot X_i + \beta_i \cdot Y_i\right) \\ &= \left(\alpha_i - \alpha_i'\right) \cdot X_i = \left(\alpha_i - \alpha_i'\right) \cdot a \cdot P.\end{aligned} \tag{5}$$

$\mathscr{C}$ outputs $(d_i - d_i') \cdot (\alpha_i - \alpha_i')^{-1}$ as the solution of the given DL problem. To compute the probability that $\mathscr{C}$ solves the DL problem, three related events are listed below.

(i) E_1: id_i equals id_I.

(ii) E_2: $\mathscr{C}$ is able to forge a legal ciphertext.

Let q_{h_1} denote the number of h_1 involved in the game. It is easy to get that $\Pr[E_1] = 1/q_{h_1}$ and $\Pr[E_2|E_1] = \epsilon$. Then, the probability that $\mathscr{C}$ solves the DL problem is $\eta = \Pr[E_1 \wedge E_2] = \Pr[E_2|E_1] \cdot \Pr[E_1] = \epsilon/q_{h_1}$. Because of the nonnegligibility of ϵ, we know that η is nonnegligible. This is in contradiction with the hardness of the DL problem. Thus, the proposed data aggregation scheme is able to provide unforgeability.

5.3. Analysis of Security Requirements. We will show that the proposed lightweight data aggregation scheme is able to meet security requirements presented in Section 3.

(i) Confidentiality. The internal attacker against the proposed data aggregation scheme can compute $c = \sum_{i=1}^{n}(c_i - h_2(s \cdot Y_i))$. Without the blinding factor c, he/she cannot extract the sum of the power consumption data by computing $m = c + \theta_0 \bmod q$. Besides, Theorem 4 shows that the proposed lightweight data aggregation scheme can supply confidentiality against any external attacker. Thus, our lightweight data aggregation scheme can supply confidentiality.

(ii) Authentication. Theorem 3 shows that any attacker cannot forge a legal ciphertext. Then, Agg can verify the legality of received messages by verifying if the equation $(\sum_{i=1}^{n} z_i \cdot d_i) \cdot P = (\sum_{i=1}^{n} z_i)P_{\text{pub}} + \sum_{i=1}^{n}(z_i \cdot \alpha_i \cdot X_i + z_i \cdot \beta_i \cdot Y_i)$ holds. Therefore, the proposed data aggregation scheme can provide authentication.

(iii) Integrity. Theorem 3 demonstrates that any attacker against the proposed data aggregation scheme cannot forge a legal ciphertext. Agg can detect any modification of the received data by verifying if the equation $(\sum_{i=1}^{n} z_i \cdot d_i) \cdot P = (\sum_{i=1}^{n} z_i)P_{\text{pub}} + \sum_{i=1}^{n}(z_i \cdot \alpha_i \cdot X_i + z_i \cdot \beta_i \cdot Y_i)$ holds. Therefore, the proposed data aggregation scheme can provide integrity.

(iv) Resistance against Attacks. The proposed lightweight data aggregation scheme can resist the replay attack, the modification attack, the man-in-the-middle attack, and the impersonation attack. The reason is analyzed below.

(1) Replay Attack. The timestamp t is involved in the ciphertext. Agg can find any reply of previous message by verifying t's freshness. Thus, the proposed lightweight data aggregation scheme can resist the replay attack.

(2) Modification Attack. Theorem 3 demonstrates that any attacker against the proposed data aggregation scheme cannot forge a legal ciphertext. Agg can detect any modification of the received data by verifying if $(\sum_{i=1}^{n} z_i \cdot d_i) \cdot P = (\sum_{i=1}^{n} z_i)P_{\text{pub}} + \sum_{i=1}^{n}(z_i \cdot \alpha_i \cdot X_i + z_i \cdot \beta_i \cdot Y_i)$ holds. Thus, the proposed lightweight data aggregation scheme can resist the modification attack.

(3) Man-in-the-Middle Attack. The above analysis demonstrates that the proposed lightweight data aggregation scheme can supply authentication; that is, Agg can authenticate SM_i by checking if $d_i \cdot P = P_{\text{pub}} + \alpha_i \cdot X_i + \beta_i \cdot Y_i$ holds. Thus, the proposed lightweight data aggregation scheme can resist the man-in-the-middle attack.

(4) Impersonation Attack. Theorem 4 shows that any attacker cannot forge a legal ciphertext without SM_i's secret key. Then, Agg can detect any impersonation by verifying the validity of the received ciphertext. Therefore, the proposed lightweight data aggregation scheme can resist the impersonation attack.

6. Performance Analysis

We analyze both computation and communication costs of our lightweight data aggregation scheme in this section. We also compare its performance with two of the most recently proposed data aggregation schemes to show its lightweight costs.

To achieve a fair comparison, we compare recently proposed aggregation schemes under the same security level. In the BGN encryption scheme [37], two 512-bit prime numbers $p = 2 \cdot p' + 1$ and $q = 2 \cdot q' + 1$ are applied in our experiments, where p' and q' are also large prime numbers. In schemes based on bilinear pairing, a Tate pairing based on a Type A elliptic curve $\widehat{E} : y^2 = x^3 + 1 \bmod \widehat{p}$ with a prime order $\widehat{q}$ is applied in our experiments, where the lengths of $\widehat{p}$ and $\widehat{q}$ are 512 bits and 160 bits, respectively. In schemes based on ECC, an elliptic curve $\overline{E} : y^2 = x^3 + a \cdot x + b \bmod \overline{p}$ with a prime order $\overline{q}$ is applied in our experiments, where the lengths of $\widehat{p}$ and $\widehat{q}$ are 160 bits.

6.1. Analysis of Computation Costs. Based on the well-known cryptographic library MIRACL [42], we have implemented all related operations on a personal computer with an Intel I5-3210M 2.50 GHz Center Processor Unit (CPU), an 8 Gbyte Random Access Memory (RAM), and the Windows 7 operation system. Table 3 presents the operations' notations and runtime results.

Each SM_i in the Fan et al.'s scheme [35] runs one BGN encryption operation, one exponentiation in BGN algorithm, two multiplications related to BGN algorithm, one HTP_{G_1} operation, one PM_{G_1} operation, and one general hash function. Therefore, SM_i's runtime is $T_{\text{ENC}_{\text{BGN}}} + T_{\text{EXP}_{\text{BGN}}} + 2 \times T_{\text{MUL}_{\text{BGN}}} + T_{\text{HTP}_{\text{BP}}} + T_{\text{PM}_{\text{BP}}} + T_{\text{GH}} = 8.315 + 8.096 + 2 \times 0.032 + 14.293 + 5.485 + 0.001 = 36.254$ ms. Agg in Fan et al.'s scheme [35] runs one BGN decryption, one exponentiation

FIGURE 3: Runtime comparisons of related schemes.

related to the BGN algorithm, $n-1$ multiplication related to BGN algorithm, n hash-to-point, $n+1$ bilinear pairing, $2n$ point multiplication related to the bilinear pairing, $n-1$ point multiplication with a short exponent related to the bilinear pairing, $n-1$ exponentiation related to the bilinear pairing, and one general hash function. Therefore, Agg's runtime is $T_{\mathrm{DEC}_{\mathrm{BNG}}} + T_{\mathrm{exp-BNG}} + (n-1)\cdot T_{\mathrm{MUL}_{\mathrm{BNG}}} + n\cdot T_{\mathrm{HTP}_{\mathrm{BP}}} + (n+1)\cdot T_{\mathrm{BP}} + (2n)\cdot T_{\mathrm{PM}_{\mathrm{BP-}s}} + (n-1)\cdot T_{\mathrm{PA}_{\mathrm{BP}}} + (n-1)\cdot T_{\mathrm{EXP}_{\mathrm{BP}}} + T_{\mathrm{GH}} = 4.056 + 8.096 + (n-1)\times 0.032 + n\times 14.293 + (n+1)\times 17.001 + (2n)\times 0.343 + (n-1)\times 0.023 + (n-1)\times 0.874 + 0.001 = (32.909\cdot n + 28.225)$ microseconds.

Each SM_i in the proposed scheme executes two point multiplication operations related to ECC and two general hash functions. Therefore, SM_i's runtime is $2\times T_{\mathrm{PM}_{\mathrm{ECC}}} + 2\times T_{\mathrm{GH}} = 2\times 0.986 + 2\times 0.001 = 1.974$ microseconds. Agg in the proposed scheme executes $3\times n+2$ point multiplication related to ECC, $2\times n$ point addition related to ECC, and $3\times n$ general hash functions. Therefore, Agg's runtime is $(3\times n+2)\times T_{\mathrm{PM}_{\mathrm{ECC}}} + 2\times n\times T_{\mathrm{PA}_{\mathrm{ECC}}} + 3\times n\times T_{\mathrm{GH}} = (3\times n+2)\times 0.986 + 2\times n\times 0.004 + 3\times n\times 0.001 = 2.969\cdot n + 1.972$.

Table 4 and Figure 3 show the runtime comparisons among Fan et al.'s data aggregation scheme [35], He et al.'s scheme [4], and the proposed scheme. From Tables 4 and 2, the proposed scheme incurs a lower computation cost as compared to Fan et al.'s scheme and He et al.'s scheme at both sides of SM_i and Agg.

6.2. Analysis of Communication Costs. Since the sizes of p_1, q_1, p', q', $\widehat{p}$, and $\widehat{q}$ are 512 bits, 512 bits, 512 bits, 160 bits, 1024 bits, and 160 bits, respectively, we can determine that the sizes of elements in Z_n^*, G_1, G_2, $Z_{q'}^*$, $Z_{\widehat{p}}^*$, and $Z_{\widehat{q}}^*$ are 1024 bits, 1024 bits, 1024 bits, 160 bits, 1024 bits, and 160 bits, respectively. We assume that the size of both the timestamp and the identity are each 32 bits. The communication costs of the related data aggregation schemes are shown below.

In Fan et al.'s data aggregation scheme [35] SM_i sends the message $(\delta_i, \mathrm{CT}_i, t)$ to Agg, where $\delta_i \in G_1$, $\mathrm{CT}_i \in Z_n^*$, and t is the timestamp. Therefore, the communication cost of Fan et al.'s data aggregation scheme is $1024 + 1024 + 32 = 2080$ bits. In He et al.'s data aggregation scheme [4] SM_i sends the message $(\mathrm{ID}_i, R_i, \delta_i, \mathrm{CT}_i, t)$ to Agg, where $R_i \in G_1$, $\delta_i \in Z_{q'}^*$, $\mathrm{CT}_i \in G_1$, ID_i is SM_i's identity, and t is the timestamp. Therefore, the communication cost of He et al.'s data aggregation scheme is $32 + 1024 + 160 + 1024 + 32 = 2272$ bits. In the proposed data aggregation scheme, SM_i sends the message (c_i, d_i, e_i, t) to Agg, where $c_i \in Z_n^*$, $d_i \in Z_{\widehat{p}}^*$, $e_i \in Z_{\widehat{q}}^*$, and t is the timestamp. Therefore, the communication cost of the proposed data aggregation scheme is $1024 + 1024 + 160 + 32 = 2240$ bits.

Based on the above evaluation, we note that the proposed data aggregation scheme incurs lower communication cost than He et al.'s data aggregation scheme. The proposed data aggregation scheme incurs a higher communication cost than Fan et al.'s data aggregation scheme. Security is most important for a data aggregation scheme. Therefore, it is reasonable to address serious security weaknesses in Fan et al.'s data aggregation scheme at the cost of increasing the communication cost slightly.

7. Conclusion

To ensure security and protect privacy in the smart grid environment, several data aggregation schemes have been proposed in recent years. However, most of them are not secure against internal attackers. To address the problem, Fan et al. [35] proposed a data aggregation scheme to mitigate internal attacks. Unfortunately, their data aggregation scheme suffers from serious security weaknesses. To enhance security, He et al. [4] proposed an improved data aggregation scheme using bilinear pairing. However, the performance of He et al.'s scheme is not very suitable for the smart grid environment because the smart meter has limited computation capability. In this paper, we have proposed a novel data aggregation scheme that can thwart internal attacks for the smart grid environment. The security analysis shows that the proposed scheme is provably secure and can meet the security requirements. Besides, performance evaluation results show that the proposed scheme incurs lower communication costs. The stronger security and better performance of the proposed scheme demonstrate that it is more suitable for smart grids.

With the fast development of quantum computing, the traditional mathematical problems (such as the DL problem and the CDH problem) are likely to be solved in polynomial time by quantum computers. Subsequently, all the above data aggregation schemes for the smart grid will not be secure at all. The lattice has been widely used to construct many cryptographic schemes that can provide resistance against the strong capabilities of quantum computers. However, no data aggregation scheme based on the lattice has been proposed yet. To improve security, it is worthwhile to consider the design of a data aggregation scheme for the smart grid based on the lattice approach.

Table 2: The aggregation phase.

SM_i		Agg
Extract m_i; generate $y_i \in Z_q^*$; $Y_i = y_i \cdot P$; $\widehat{Y}_i = y_i \cdot P_{\text{pub}}$; $c_i = m_i + \theta_i + h_2(\widehat{Y}_i) \bmod q$; $\beta_i = h_3(\text{id}_i, X_i, Y_i, c_i, t)$; $d_i = s_i + \beta_i \cdot y_i \bmod q$	$\xrightarrow{\{X_i, Y_i, c_i, d_i, t\}}$	$\alpha_i = h_1(\text{id}_i, X_i)$; $\beta_i = h_3(\text{id}_i, X_i, Y_i, c_i, t)$; generate $z_1, \ldots, z_n \in [1, 2^w]$; check $\left(\sum_{i=1}^{n} z_i \cdot d_i\right) \cdot P \stackrel{?}{=} \left(\sum_{i=1}^{n} z_i\right) \cdot P_{\text{pub}} + \sum_{i=1}^{n} (z_i \cdot \alpha_i \cdot X_i + z_i \cdot \beta_i \cdot Y_i)$; $c = \sum_{i=1}^{n} (c_i - h_2(s \cdot Y_i))$; $m = c - \theta_0$

Table 3: Notations about related operations and runtime results (microseconds).

Notation	Description	Runtime
ENC_{BGN}	BGN encryption	8.315
DEC_{BGN}	BGN decryption	4.056
EXP_{BGN}	Exponentiation related to BGN algorithm	8.096
MUL_{BGN}	Multiplication related to BGN algorithm	0.032
BP	Bilinear pairing	17.001
HTP	Hash-to-point	14.293
PM_{BP}	Point multiplication related to the bilinear pairing	5.485
$\text{PM}_{\text{BP}-s}$	Point multiplication with a short exponent related to the bilinear pairing	0.343
PA_{BP}	Point addition related to the bilinear pairing	0.023
EXP_{BP}	Exponentiation related to the bilinear pairing	0.874
MUL_{BP}	Multiplication related to the bilinear pairing	0.005
EXP_{DLP}	Exponentiation related to the DL problem	1.295
$\text{EXP}_{\text{DLP}-s}$	Exponentiation with a short exponent related to the DL problem	0.081
MUL_{DLP}	Multiplication related to the DL problem	0.012
PM_{ECC}	Point multiplication related to ECC	0.986
$\text{PM}_{\text{ECC}-s}$	Point multiplication with a short exponent related to ECC	0.061
PA_{ECC}	Point addition related to ECC	0.004
GH	General hash function	0.001

Table 4: Runtime comparisons (microseconds).

	Fan et al.'s scheme	He et al.'s scheme	The proposed scheme
SM_i	36.254	20.145	**17.751**
Agg	$32.909n + 28.225$	$6.264n + 48.249$	$\mathbf{2.969n + 1.972}$

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (nos. 61572370, 61501333, 61572379, and U1536204), the National High-Tech Research and Development Program of China (863 Program) (no. 2015AA016004), and the Natural Science Foundation of Hubei Province of China (no. 2015CFB257). Sherali Zeadally's work has been supported by a University Research Professorship Award from the University of Kentucky.

References

[1] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[2] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive machine-to-machine communications: visions and potentials for the smart grid," *IEEE Network*, vol. 26, no. 3, pp. 6–13, 2012.

[3] C. Greer, D. A. Wollman, D. E. Prochaska et al., "Nist framework and roadmap for smart grid interoperability standards, release 3.0".

[4] D. He, N. Kumar, and J.-H. Lee, "Privacy-preserving data aggregation scheme against internal attackers in smart grids," *Wireless Networks*, vol. 22, no. 2, pp. 491–502, 2016.

[5] J. Gao, Y. Xiao, J. Liu, W. Liang, and C. L. P. Chen, "A survey of communication/networking in smart grids," *Future Generation Computer Systems*, vol. 28, no. 2, pp. 391–404, 2012.

[6] N. Saputro, K. Akkaya, and S. Uludag, "A survey of routing protocols for smart grid communications," *Computer Networks*, vol. 56, no. 11, pp. 2741–2771, 2012.

[7] V. C. Güngör, D. Sahin, T. Kocak et al., "Smart grid technologies: communication technologies and standards," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 4, pp. 529–539, 2011.

[8] W. Su, H. Eichi, W. Zeng, and M.-Y. Chow, "A survey on the electrification of transportation in a smart grid environment," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 1–10, 2012.

[9] D. Wu and C. Zhou, "Fault-tolerant and scalable key management for smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 375–381, 2011.

[10] Z. Wan, G. Wang, Y. Yang, and S. Shi, "SKM: scalable key management for advanced metering infrastructure in smart grids," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 12, pp. 7055–7066, 2014.

[11] K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang, and T. Sato, "A key management scheme for secure communications of information centric advanced metering infrastructure in smart grid," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2072–2085, 2015.

[12] J. H. Park, M. Kim, and D. Kwon, "Security weakness in the smart grid key distribution scheme proposed by xia and wang," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1613–1614, 2013.

[13] J.-L. Tsai and N.-W. Lo, "Secure Anonymous Key Distribution Scheme for Smart Grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 906–914, 2016.

[14] D. He, H. Wang, M. K. Khan, and L. Wang, "Lightweight anonymous key distribution scheme for smart grid using elliptic curve cryptography," *IET Communications*, vol. 10, no. 14, pp. 1795–1802, 2016.

[15] H. Liu, H. Ning, Y. Zhang, and M. Guizani, "Battery status-aware authentication scheme for V2G networks in smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 99–110, 2013.

[16] H. Li, R. Lu, L. Zhou, B. Yang, and X. Shen, "An efficient Merkle-tree-based authentication scheme for smart grid," *IEEE Systems Journal*, vol. 8, no. 2, pp. 655–663, 2014.

[17] N. Saxena, B. J. Choi, and R. Lu, "Authentication and Authorization Scheme for Various User Roles and Devices in Smart Grid," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 907–921, 2016.

[18] N. Koblitz, "Elliptic curve cryptosystems," *Mathematics of Computation*, vol. 48, no. 177, pp. 203–209, 1987.

[19] V. S. Miller, "Use of elliptic curves in cryptography," in *Proceedings of the Conference on the Theory and Application of Cryptographic Techniques*, pp. 417–426, Springer, 1985.

[20] X. Huang, Y. Xiang, E. Bertino, J. Zhou, and L. Xu, "Robust multi-factor authentication for fragile communications," *IEEE Transactions on Dependable and Secure Computing*, vol. 11, no. 6, pp. 568–581, 2014.

[21] X. Huang, Y. Xiang, A. Chonka, J. Zhou, and R. H. Deng, "A generic framework for three-factor authentication: Preserving security and privacy in distributed systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1390–1397, 2011.

[22] D. He, S. Zeadally, N. Kumar, and J.-H. Lee, "Anonymous authentication for wireless body area networks with provable security," *IEEE Systems Journal*, 2016.

[23] Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, "Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing," *IEICE Transactions on Communications*, vol. E98B, no. 1, pp. 190–200, 2015.

[24] Z. Fu, X. Wu, C. Guan, X. Sun, and K. Ren, "Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2706–2716, 2016.

[25] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," *IEEE Transactions on Parallel and Distributed Systems*, vol. 9, no. 27, pp. 2546–2559, 2015.

[26] Z. Xia, X. Wang, X. Sun, Q. Liu, and Q. Wang, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," *IEEE Transactions on Parallel and Distributed Systems*, vol. 2, no. 27, pp. 340–352, 2015.

[27] X. F. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations," *IEEE Transactions On Parallel And Distributed Systems*, vol. 25, no. 9, pp. 2386–2396, 2014.

[28] X. Chen, J. Li, X. Huang, J. Ma, and W. Lou, "New Publicly Verifiable Databases with Efficient Updates," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 546–556, 2015.

[29] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption,," in *Proceedings of the 1st IEEE International Conference on Smart Grid Communications (SmartGridComm '10)*, pp. 327–332, 2010.

[30] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Preceedings of the International Workshop on Security and Trust Management*, pp. 226–238, Springer, 2010.

[31] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," vol. 1592, pp. 223–238, Springer, Berlin, Germany, Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques.

[32] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1632, 2012.

[33] J. Zhang, L. Liu, Y. Cui, and Z. Chen, "SP^{2}DAS: self-certified PKC-based privacy-preserving data aggregation scheme in smart grid," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 457325, 11 pages, 2013.

[34] L. Chen, R. Lu, and Z. Cao, "PDAFT: a privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 8, no. 6, pp. 1122–1132, 2014.

[35] C.-I. Fan, S.-Y. Huang, and Y.-L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666–675, 2014.

[36] H. Bao and R. Lu, "Comment on 'Privacy-Enhanced Data Aggregation Scheme Against Internal Attackers in Smart Grid'," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 2–5, 2016.

[37] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-dnf formulas on ciphertexts," in *Proceedings of the Theory of Cryptography Conference*, vol. 3378 of *Lecture Notes in Comput. Sci.*, pp. 325–341, Springer, Berlin, Gremany, 2005.

[38] H. Cohen, G. Frey, R. Avanzi et al., *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, CRC Press, 2005.

[39] J. K. Liu, T. H. Yuen, M. H. Au, and W. Susilo, "Improvements on an authentication scheme for vehicular sensor networks," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2559–2564, 2014.

[40] M. Barbosa and P. Farshim, "Certificateless signcryption," in *Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS '08)*, pp. 369–372, ACM, March 2008.

[41] D. Pointcheval and J. Stern, "Security arguments for digital signatures and blind signatures," *Journal of Cryptology*, vol. 13, no. 3, pp. 361–396, 2000.

[42] M. Scott, Miracl library, 2011, http://www.shamus.ie.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 9731806, 10 pages
https://doi.org/10.1155/2017/9731806

Research Article

Semantic Interoperability in Heterogeneous IoT Infrastructure for Healthcare

Sohail Jabbar,[1] Farhan Ullah,[2] Shehzad Khalid,[3] Murad Khan,[1] and Kijun Han[1]

[1]*School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea*
[2]*Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal, Pakistan*
[3]*Department of Computer Engineering, Bahria University, Islamabad, Pakistan*

Correspondence should be addressed to Kijun Han; kjhan@knu.ac.kr

Received 21 December 2016; Accepted 6 February 2017; Published 5 March 2017

Academic Editor: Yoshikazu Miyanaga

Interoperability remains a significant burden to the developers of Internet of Things' Systems. This is due to the fact that the IoT devices are highly heterogeneous in terms of underlying communication protocols, data formats, and technologies. Secondly due to lack of worldwide acceptable standards, interoperability tools remain limited. In this paper, we proposed an IoT based Semantic Interoperability Model (IoT-SIM) to provide Semantic Interoperability among heterogeneous IoT devices in healthcare domain. Physicians communicate their patients with heterogeneous IoT devices to monitor their current health status. Information between physician and patient is semantically annotated and communicated in a meaningful way. A lightweight model for semantic annotation of data using heterogeneous devices in IoT is proposed to provide annotations for data. Resource Description Framework (RDF) is a semantic web framework that is used to relate things using triples to make it semantically meaningful. RDF annotated patients' data has made it semantically interoperable. SPARQL query is used to extract records from RDF graph. For simulation of system, we used Tableau, Gruff-6.2.0, and Mysql tools.

1. Introduction

Internet of things (IoT) is a combination of spatially distributed smart objects which have sensing capabilities and embedded identification through RFID technology. Specifically, the integration of sensors, RFID tags, and communicating technologies forms the underpinning of IoT. It addresses the traceability, visibility, and controllability of smart objects. It is a future vision through which digital and physical objects can be interlinked and intercommunicated to provide some domain specific services. IoT transforms real world objects into smart objects which can sense the environmental physical quantity and communicate it accordingly. RFID technology is now widely used in tracking of people, objects, and animals. Electronic Product Codes are encoded in RFID tags which can be used to track smart objects in IoT [1, 2]. For storage of incoming data from these smart devices, Cloud technology is one of the suitable options. There is a long list of its applications like environmental monitoring [3], healthcare service [4], inventory and production management [5], food supply chain [6], transportation and logistics [7], smart cities [8], smart homes [9], data collection through aerial vehicles [10], firefighting system [11], social networks [12], and VANETs [13] just to name a few.

In spite of its wide spread emergence, IoT is still in its infant stage and has big room for research in variety of issues like standards, scalability, heterogeneity of different devices, common service description language, domain specific service discovery, integration with existing IT systems, and so forth. In IoT, one vital issue is interoperability among smart objects that is the ability to interconnect and communicate different vendors' systems to form a cost effective and easy to implement network. To promote enterprise interoperability, C4ISR architecture working group (AWG) developed the LISI (Levels of Information Systems Interoperability) approach in 1997 [14]. The purpose of LISI model is to provide maturity model to US Department of Defense (DoD) and to specify interoperability needs and selecting pragmatic solutions for enterprise interoperability [15]. IoT industry is working to migrate the interoperability standards to existing

IP and Ethernet standards to work with existing infrastructure [16]. According to Cisco analysis, there will be 50 billion connected devices [17]. The IoT Forums from worldwide are working to develop a common model that can ensure interoperability among smart objects. Open Internet Consortium (OIC) is currently focusing on the IoT interoperability to define specifications, integration of billions of smart objects, and scalability issues (http://blogs.cisco.com/digital/iot-meets-standards-driving-interoperability-and-adoption). Low cost interoperability among smart objects is an important factor for smart cities. Smart mesh backbone gateways (WRT54GL by Linksys (http://www.linksys.com/us/support-product?pid=01t80000003KOkNAAW) and net5501 by Soekris Engineering Inc. (http://soekris.com/products/eol-products/net5501.html)) are developed for smart cities which provide low cost interoperability. The Grid-Wise Architecture Council (GWAC) mission is to enable interoperability among the objects that interact with the electric power system. The GWAC introduced a context setting framework which identifies interoperability issues [1]. The solution of interoperability will help the customers to continue working in mixed environments in nearer future. It will solve the complexity of an organization's infrastructure, reduce the cost of buildings, and help in supporting heterogeneous infrastructure. Energy aware algorithms will be required to ask from users for routing their packets on wireless or IEEE 802.11 mesh network. Thus, it will save the energy of smart objects [18]. Users with multiple vendors' technologies from worldwide bring scalability in IoT. It will increase the workflow efficiency in any environment. It will connect objects from anywhere to anywhere on different technologies.

A productive effort in developing widely accepted standards can smoothly lead to solving the interoperability issue. The development of standards is for ease of plug and play device interaction, reconfigurability, and information exchange. All smart objects from multiple vendors should operate on common standards. If we can integrate IoT with existing IT infrastructure then it will solve many problems like protocol, packet size, encryption/decryption algorithms, and so forth. Operating smart objects to existing Internet Protocol version 6 (IPv6) structure may be costly because it is difficult to embed RJ45/WLAN port in every smart object but IoT industry is currently working to migrate the IoT standards to existing IPv6 and Ethernet standards to ensure the interoperability with legacy protocols. IP for Smart Objects (IPSO) (http://www.ipso-alliance.org) promotes IPv6 as a unique identification for discovering, accessing, and communicating smart objects [15]. Secondly in case of device roaming, a specific device identity is required and not IP address. Therefore every device must need a unique identification worldwide.

Collaboration among different vendors can also be a workable effort towards solving interoperability issues. Vendors should collaborate on these issues and develop a transparent method for interoperability that can be implemented in every domain. Collaboration is needed between enterprise modeling communities in developing business oriented architecture and software engineering companies working on IT oriented solutions [15]. Interoperability methods should not be domain specific. Interoperability can be solved if communicating smart objects are semantically interoperable [19]. Semantics brings what data is transferring. It gives the details about structure of transferred data. IoT device's semantics details, interpretation, and exchange of information must be developed in order to remove semantics conflicts in interoperability. The meaning responding from a smart object is fully interpretable as originally means by the smart object [2, 20]. Strong infrastructure is needed for interoperability to fit any device from multiple vendors. Consumer should feel free to use any device from any vendor and vendor should take care of consumer trust. Interoperability has a major issue of global heterogeneity because IoT devices are highly heterogeneous. A huge number of IoT devices from unknown vendors have different configuration and installation methods. Devices from multiple vendors have different methods for semantics and syntactic interoperability methods, so there will be semantic and syntactic conflicts. It will be difficult to add a new device in IoT network without resolving semantic ambiguity. In healthcare domain it can provide high quality care to patients when every device works in "interoperable environment." In healthcare industry it will automate the hospitals that would monitor the patients by physicians remotely. Physicians could get information about patients quickly and advise them for treatment. It will provide fast access to patients' records from different IT storage systems located anywhere.

In this research work, Resource Description Framework is used to provide Semantic Interoperability among physicians and patients using heterogeneous IoT devices. We have analyzed dataset on proposed IoT based Semantic Interoperability Model to provide interoperability in IoT devices. Patients' dataset semantically annotated to convert it to RDF and SPARQL query is used to extract data of any patient at any time. In proposed model medicines' data of pharma companies stored on Cloud and Physicians can access information due to semantics interoperability. Collected data is transformed into linked data (RDF triples) and SPARQL queries are applied to extract patient related semantic information from that linked data. A similar work is done by Malik et al. in [21]. A lightweight model for semantic annotation of data using heterogeneous devices in IoT has been proposed to make IoT devices interoperable. Communicating data between physician and patient is semantically annotated which solve the issue of interoperability among IoT devices.

This paper is further divided into 5 sections: Literature Review, Proposed Scheme: IoT Based Semantic Interoperability Model (IoT-SIM), Semantic Annotations of Data Using Heterogeneous IoT Devices, Experiments and Results, and Conclusion and Future Work.

2. Literature Review

IoT devices from different vendors can be integrated in healthcare environment to assist Physicians and patients remotely. Physicians can monitor their patients anytime and anywhere and can update prescription when needed. In [22], the authors used Intelligent Personal Assistant (IPA)

as software agent in IoT device for physician to give real time information about the monitored patients. The AMBRO mobile gateway collects information from different IPA devices and then take action(s). It enabled interoperability among different IPA devices. It uses only specific operating system and does not integrate with any other operating system such as iOS or windows phone. In order to communicate large number of smart devices with heterogeneous capabilities, semantic web technologies are promising tools for this purpose to share data and exchange their services efficiently. In [23], the authors proposed a semantic model for description of smart objects using ontologies and description logics to enable Semantic Interoperability. This semantic model still needs further empirical evaluation to enhance the service classifications for Semantic Interoperability among smart objects. To digitize agriculture domain, there is a need to collect information about crop growth monitoring and irrigation decisions support through smart objects.

Jayaraman et al. [24] described an OpenIoT platform used for digital agriculture use case (Phenonet). To enable Semantic Interoperability, OpenIoT platform used ontologies to represent Phenonet domain concepts to collect smart collection of information, annotation, and validation processes. A scalable and intelligent IoT architecture is required for future era to enable discovery of physical sensors and their interoperation semantically and syntactically. A semantic web enabled architecture is proposed by Desai et al. [25] to provide interoperability among smart things. The Semantic Gateway as Service (SGS) integrated semantic web technologies to enable communication between protocols such as XMPP, CoAP, and MQTT. The ontologies are used for semantic reasoning to provide Semantic Interoperability among communicating messages. Although there are different approaches developed for semantic interoperability for IoT devices, there is still lack of formal methodologies for interoperability in technology and standard format of data.

Gyrard and Serrano [26] proposed SEG 3.0 methodology to unify, federate, and provide Semantic Interoperability in IoT domain. The SEG 3.0 emerged from ontology engineering and its main benefit to integrate heterogeneous data collected from different smart things. The authors applied SEG 3.0 methodology on three different use cases, in M3 framework to support developers to design semantic-based IoT applications, the VITAL EU project for smart cities, and the FIESTA-IoT EU project for Semantic Interoperability. For satisfying level of performance among low-power heterogeneous networks, there is a need of interoperability protocols and standards. To bridge this gap, the authors in [27] described features of different protocol stacks such as Bluetooth, Bluetooth Low Energy, IEEE 802.15.4, ZigBee, 6LoWPAN, and IEEE 802.15.6. They proposed the designing of generic protocol stack that communicates with multiple radios and with different protocols simultaneously, regardless of IP-based or non IP-based networks. Mingozzi et al. [28] proposed a novel uniform platform designed to include context-awareness functionalities and showed how such functions can be exploited to automate search and selection of things through natural language. It is shown that context can be used on *things services* to extract knowledge through semantic reasoning in smart homes. Today's smart phones have vast connectivity and sensing capabilities for serving body area networks and physiologic sensors. To enable interoperability between Machine-to-Machine (M2M) and IoT, common standards are the driving needs. In this context, a design of ETSI M2M gateway integrated with libraries is proposed by Pereira et al. [29]. This is to ease the deployment of IoT applications with reduced development costs. Its performance is shown by measuring the smartphone's CPU and memory usage and battery life. In [30], authors proposed an architecture for data processing in IoT Cloud environment that supports semantics interoperability. Google Cloud and Microsoft Azure were used as multi-Cloud environment with OpenIoT architecture. Another IoT concept for digitizing the factory is proposed in [31]. Data communicates from heterogeneous IT environment to ensure data consistency. Author suggestion is the integration of IoT and PLM platforms using semantic web technologies and Open Services for Lifecycle Collaboration (OSLC) standard on tool interoperability.

3. Proposed Scheme: IoT Based Semantic Interoperability Model

In our proposed model we focused on tracking and monitoring of human diseases with respect to prescribed medicine in healthcare domain. User Interface (UI), Semantic Interoperability (SI), and Cloud Services (CS) are three major constituents of IoT-SIM. In User Interface, doctor and patient interact with each other with the help of IoT devices. Doctor can monitor and prescribe patients remotely anywhere, anytime, and without any constraint of specific vendors' device. Semantic Interoperability directly interacts with User Interface section. Interoperability among IoT devices from different vendors is a big issue. Semantic Interoperability is the exchange of information with meaningful and understandable meanings. It includes semantics in data by adding self-described information packages. Semantic Interoperability is used to ensure that IoT devices from different vendors are interoperable. IoT devices take data from UI and then add semantic annotations from semantic section to make it meaningful with shared vocabulary. Data analytics is a data science technique which draws meaningful conclusions from raw data. It brings cost reduction and better and fast decision making. Data analytics technique is applied on collected data from IoT devices. Then, it underwent semantic annotations to make it more meaningful and cost effective. The proposed architectural model for semantic interoperability in IoT is shown in Figure 1.

3.1. Semantic Annotations of Data Using Heterogeneous IoT Devices. IoT devices are communicating through sensors. Every device has sensor network API which is used to filter data according to the domain. Then this filtered data is sent to sensors' web service through which sensors are communicating with the whole world. To provide web services using heterogenous IoT devices Sensor Web Enablement framework has been used. SWE used to discover, access, and use IoT devices. These keywords are tokens which represent healthcare data related to human diseases. Then dataset is

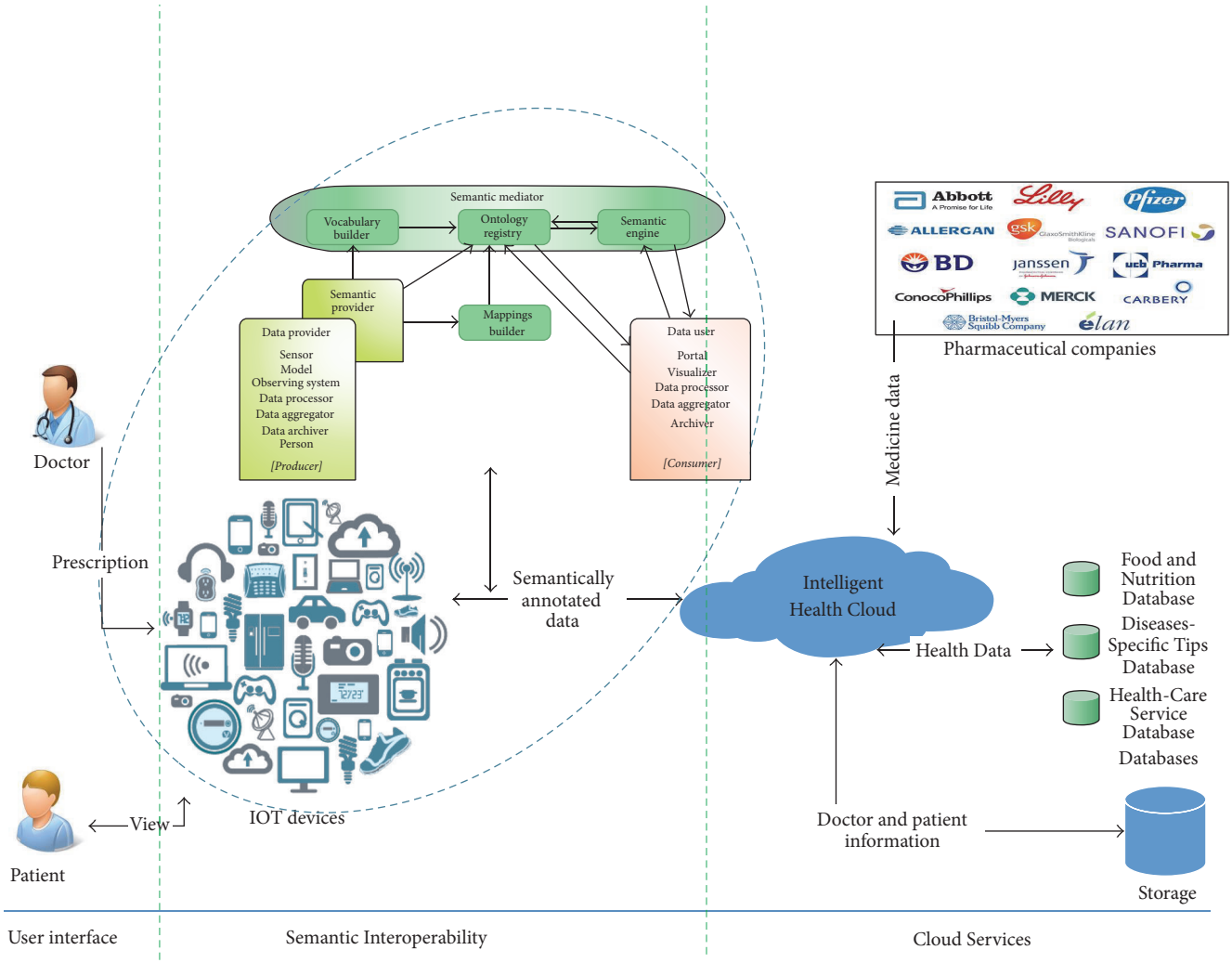


Figure 1: Architectural model for Semantic Interoperability in IoT.

moved to sematic interoperability section where every token is categorized in specific domain of hum diseases. Every token with its description is sent to diseases classification section. In this section human disease is classified to show which patient has which type of disease. A lightweight model for semantic annotation of data using heterogeneous devices in IoT that is depicted in Figure 2 represents this operational flow from data collection to classification of patients according to their diseases.

On identification of disease, system can automatically suggest medicine for it. If prescribed medicine from doctor to patient is matched with identified disease then it is correct medicine otherwise doctor prescribed the wrong medicine. Storage attached with Intelligent Health Cloud stores information about correct or wrong medicine with doctors' and patient's identification. Classified diseases in different categories according to healthcare domain are sent to Tagging section where diseases are semantically annotated automatically or manually through Resource Description Framework. RDF is a standard metadata model that uses URIs (Uniform Resource Identifier) to relate things semantically. The same is also shown in Figure 2.

4. Experiments and Results

We have analyzed our experiments on Heart Disease Dataset (Heart Disease Dataset from Center for Machine Learning and Intelligent Systems: https://archive.ics.uci.edu/ml/datasets/Heart+Disease). This dataset is collected from University Hospital, Zurich, Switzerland (switzerland.data). The principal investigator responsible for the data collection is VA Medical Center, Long Beach and Cleveland Clinic Foundation, Robert Detrano, M.D., Ph.D. The dataset contains 123 patients with different types of heart diseases. Dataset is plotted graphically with detailed information as shown in Figure 3. This graph shows that dataset consists of patient Id, age, sex, and disease type. For sex attribute, 1 denotes male and 0 denotes female while for all other attributes, 1 denotes yes and 0 denotes not. There are 14 different attributes used for a patient in a database. Dig attribute describes *Digitalis* used during exercise and its value is 1 or 0.

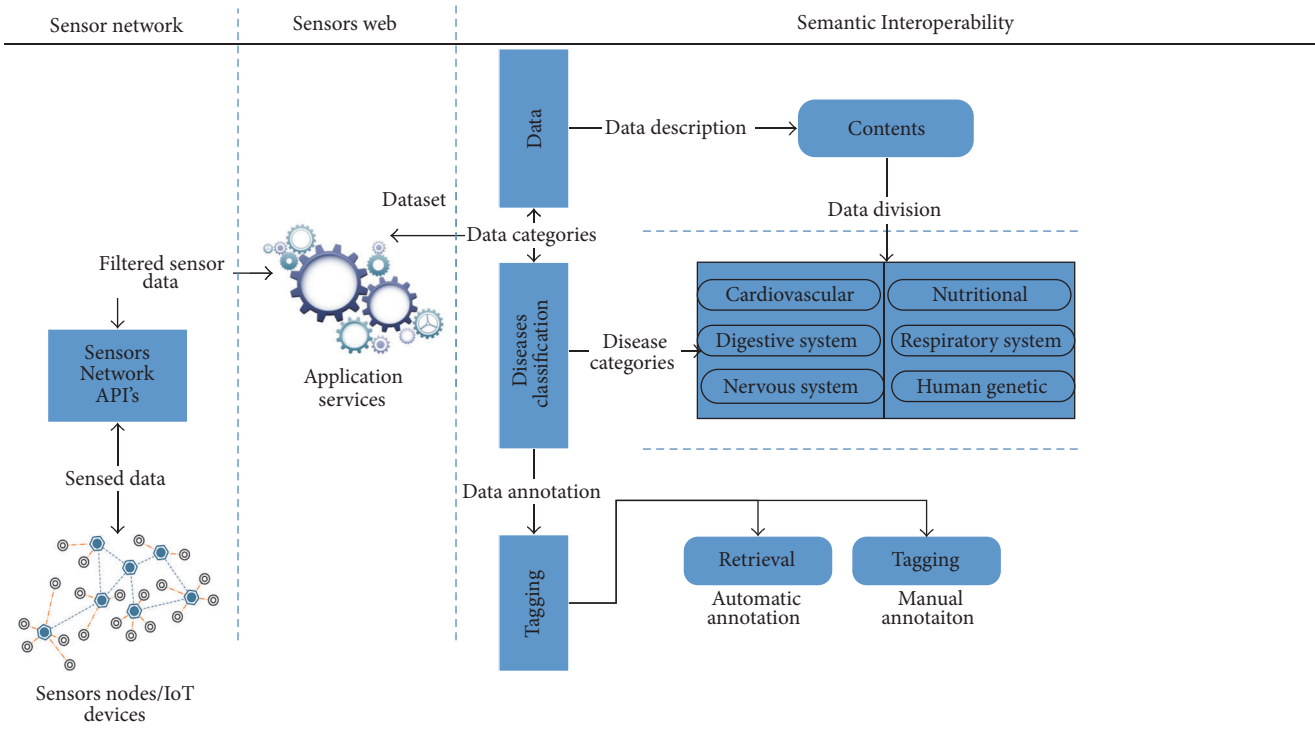


Figure 2: Lightweight Model for Semantic annotation of data using heterogeneous devices in IoT.

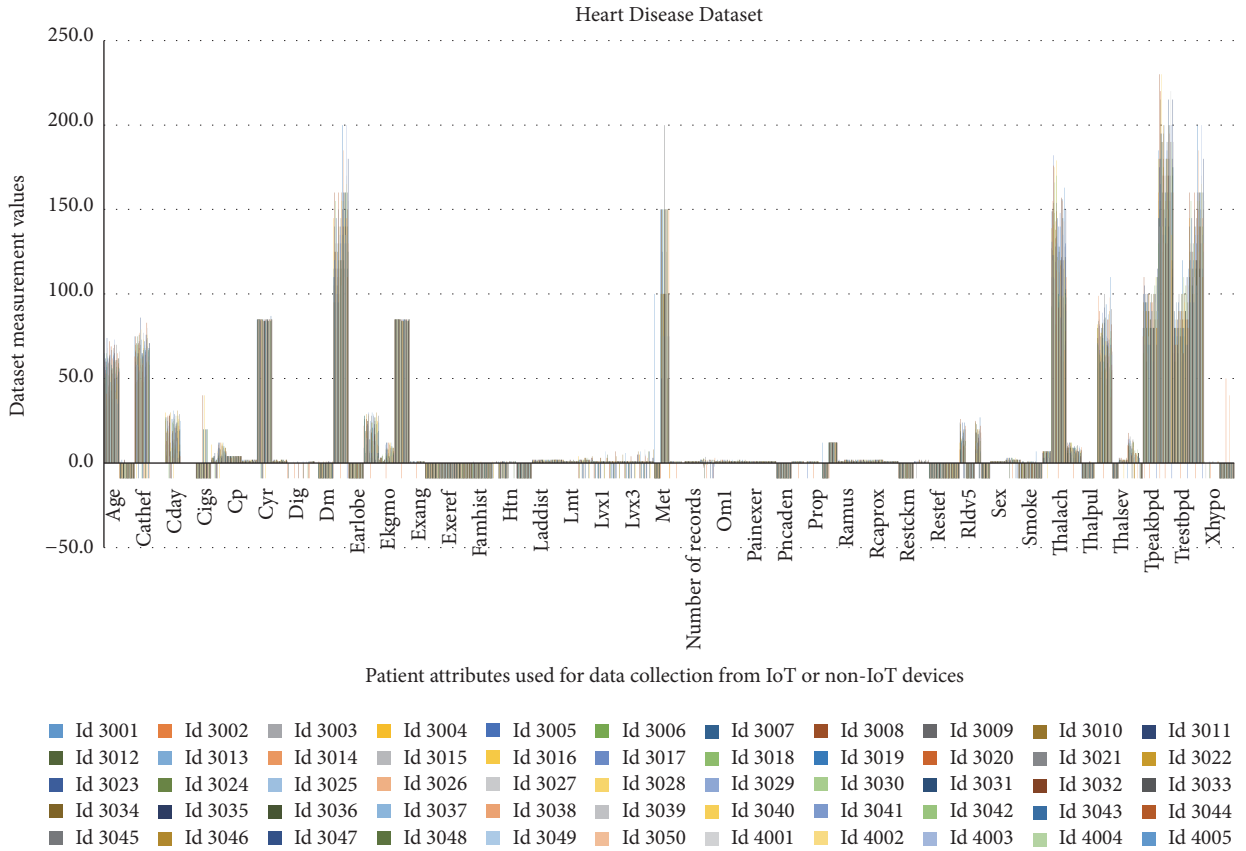


Figure 3: Patient Id-Wise dataset graph representation.

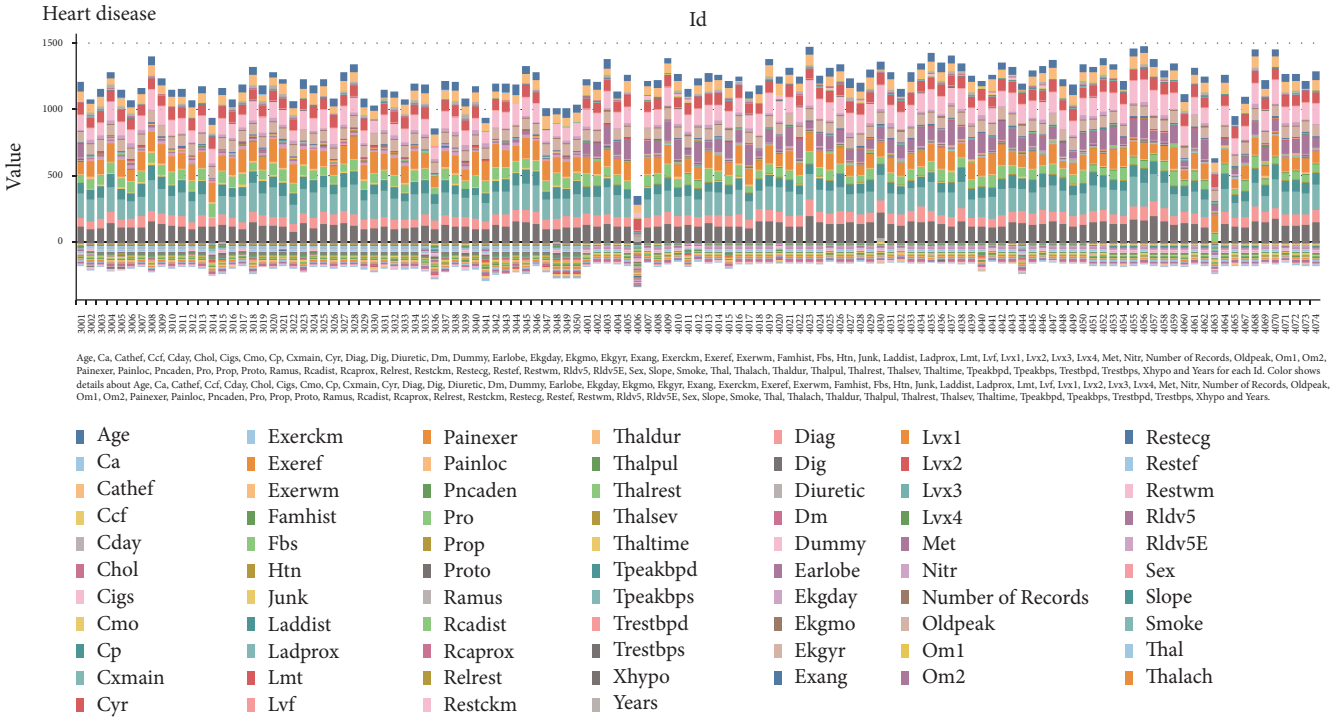


Figure 4: Accumulated heart diagnostic measurements.

All attributes are plotted horizontally and its corresponding weighted value is plotted vertically. Mostly patients are 60 to 65 years of age as shown in age attribute. In graph Tpeakbps has highest value meaning that almost all heart patients have blood pressure disease as well. Some attributes are negative in the graph. It intuits that these attributes are not used during experiment as shown in Figure 3. We used Tableau tool (http://www.tableau.com/products/desktop) to plot patients' accumulated graph against their diseases. The patients identified by their Ids are plotted at x-axis and accumulated values of their diseases are plotted at x-axis. Patients' attributes are used for data collected from IoT or non-IoT devices as shown in Figure 3.

Accumulated heart diagnostic measurements of dataset are plotted as shown in Figure 4. Figure 4 is the complete reflection of dataset using accumulated measurements of attributes in dataset which showed that the collected dataset is fully correct and validate. There are different colors used to identify diseases. Many attributes are negative of 4006 Id patient which shows that mostly diseases are not analyzed during experiment of this patient. Tpeakbps value shows in a graph almost all patients' diagnosis blood pressure disease as well. Some patients like 3014, 4006, and 4063 have Tpeakbps value in negative shows that blood pressure is not analyzed during experiment. Patients with 4023 and 4056 have diagnosis of many diseases and have almost the same number of diseases. Cigs defines cigarettes per day and its value is negative means that it is not considered during experiment. The names and social security numbers of the patients were recently removed from the database, replaced with dummy values. In this paper we have simulated our proposed IoT-SIM to provide Semantic Interoperability among heterogeneous IoT devices through RDF to make it readable by machine and human. In [32], RDF is graphical database which assign Universal Resource Identifier (URI) for identification. Further author used ontologies to identify the same entity with different meanings. We annotated semantically all the data to make triples which shows relationship among every instance. Every instance in triples is identified by some URI and every instance relates with other instances using some semantic relation.

In Figure 5 the RDF graph is simulated from patients' database in Gruff-6.2.0 tool (http://franz.com/agraph/gruff/release-history.lhtml). Every attribute has specific color for its identification. Info is a table which relates with every instance semantically and all Info nodes in RDF graph relate to only one Info table. All fields are semantically annotated using RDF triples, which shows the relation of every instance with other related instances. Patients' records against diagnostics diseases are semantically readable by machine to ensure interoperability among heterogeneous IoT devices. In [33], SPARQL, a structural query language proposed by W3C, is designed for querying RDF data. Since SPARQL queries can be represented as query graphs, a SPARQL query can be answered by performing the graph pattern matching over RDF graphs. Figure 5's different visible chunks are shown in Figure 6 which shows single node RDF graph interaction with other node. A single node interaction with different attributes in dataset using RDF graph is shown in Figure 6. All detailed information of all patients is semantically annotated using RDF graph to provide matching of information semantically ignoring the vendors' hardware details. Further to validated

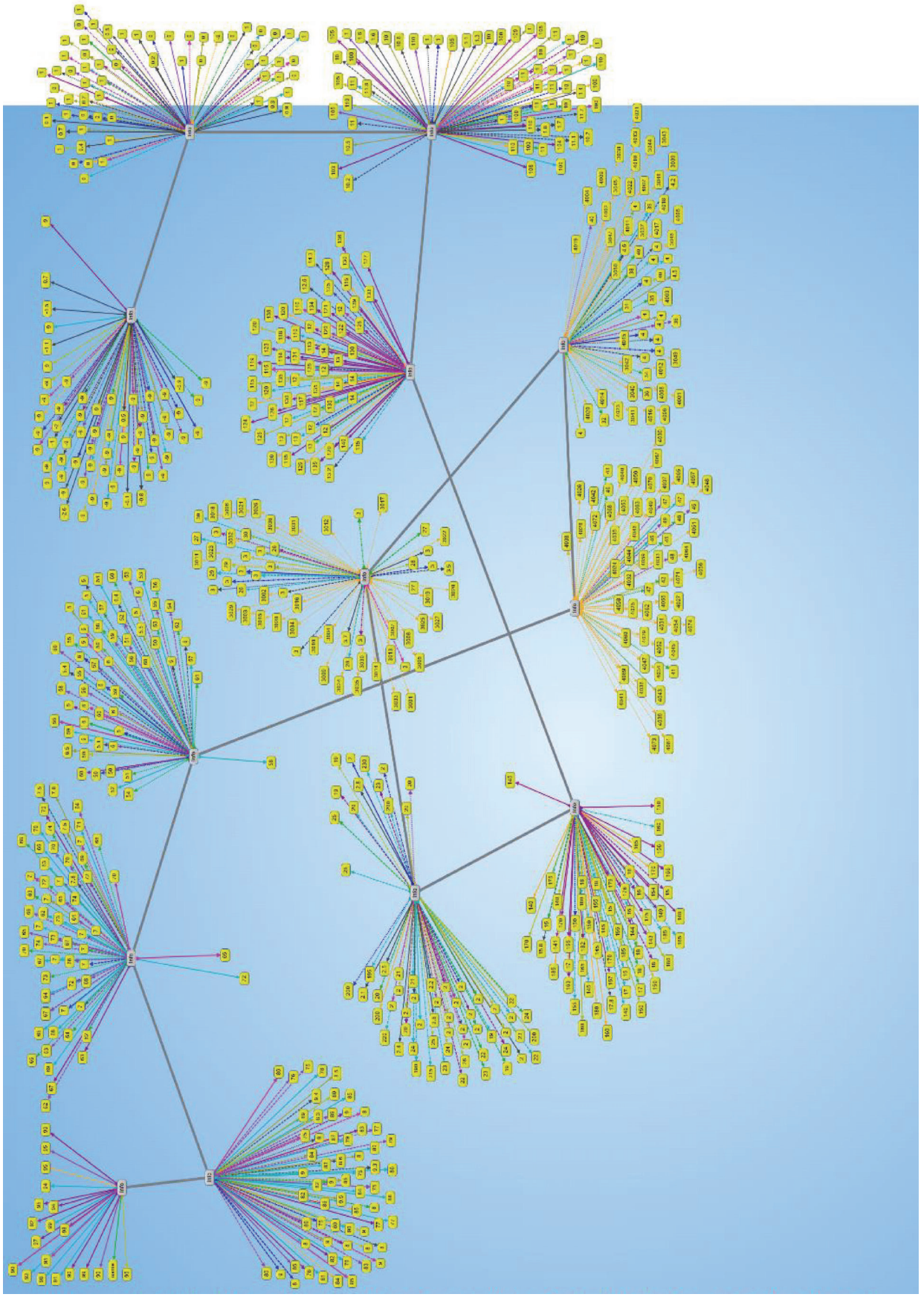
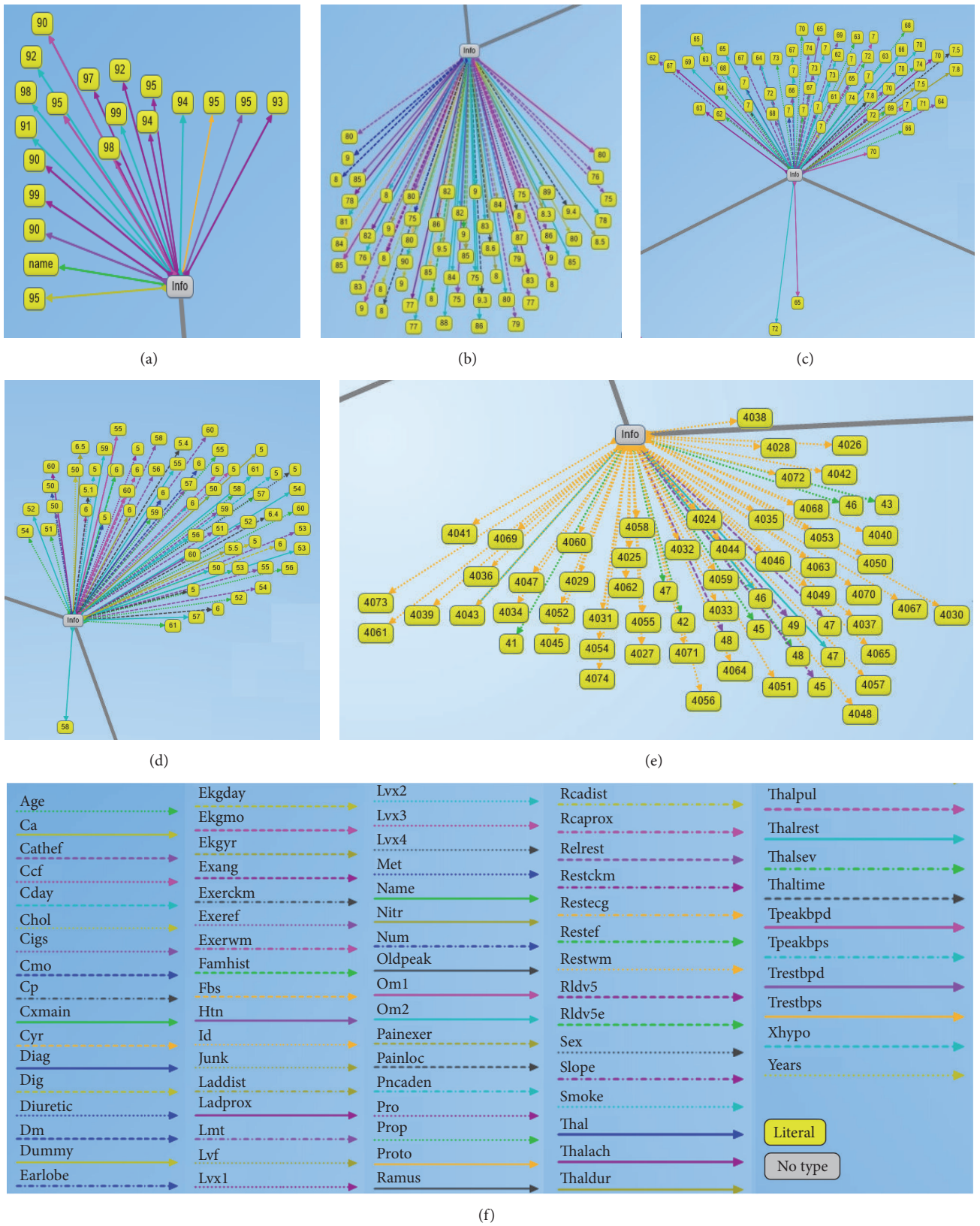


Figure 5: RDF graph for Semantic Interoperability in IoT.



(a) (b) (c)

(d) (e)

(f)

Figure 6: RDF graph chunks for Semantic Interoperability in IoT.

our experiment, we extracted records of different patients from RDF graph using SPARQL query in a text format.

The SPARQL query has been used to extract all patients records from RDF graph as RDF is semantically machine readable. We can also extract information of any patient using SPARQL query.

```
select ?var1 ?var2 ?predicateVar
where
{
?var1 ?predicateVar ?var2.
}
```

The visible chunks of simulated graph in Figure 5 are shown in Figure 6.

The attributes/legends of patients' database in RDF graph are shown in Figure 6(f). Every attribute has different color to identify in RDF graph. For example, age attribute/legend has light green color which shows every patient's age in RDF graph. These legends describe attributes in dataset which relates with patients' nodes in terms of semantically annotating.

5. Conclusion and Future Work

In this paper, we have proposed an IoT-SIM model for Semantic Interoperability among heterogeneous IoT devices in healthcare domain. The main goal of this model is to provide interoperability among heterogenous IoT devices by using semantically annotated. Physicians monitor their patients remotely anywhere, anytime, and without any constraint of specific vendors' device. RDF is used to present patients' raw data into useful information. Physicians prescribed patients against the diagnosis diseases using IoT device; then this information is semantically annotated using RDF. Further lightweight model for semantic annotation of data using heterogeneous devices in IoT has been proposed which is describing the communication among heterogenous IoT devices. To provide web services, sensors' devices communicated using SWE framework with other devices. To solve the issue of interoperability among heterogeneous IoT devices, collected dataset has been mapped to RDF graph database. The communicated data is being analyzed and annotated in terms of semantics. Annotated information is then sent to Intelligent Health Cloud where prescribed medicines match from pharmaceutical companies, then prescribed information with medicine details is sent to patient IoT. RDF graph represents patients' diseases database in triples which is semantically readable using SPARQL query. Physicians can query from his IoT device any time about the current situation of the patient from database remotely. End users do not care for time, distance, and hardware.

Our proposed model can be enhanced in future to provide syntactic interoperability among heterogeneous IoT devices. Syntactic interoperability provides the syntactic structure of the exchanged information. Next security is a hot issue in interoperability among heterogeneous IoT devices. It must be critical aspect for the solution of interoperability. The syntactic interoperability and security issues will be investigated in the future of this work.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03933566). This study was supported by the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

References

[1] S. L. Hamilton, E. W. Gunther, R. V. Drummond, and S. E. Widergren, "Interoperability—a key element for the grid and der of the future," in *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, pp. 927–931, Dallas, Tex, USA, May 2006.

[2] G. Xiao, J. Guo, L. D. Xu, and Z. Gong, "User interoperability with heterogeneous IoT devices through transformation," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1486–1496, 2014.

[3] D. Pavithra and R. Balakrishnan, "IoT based monitoring and control system for home automation," in *Proceedings of the Global Conference on Communication Technologies (GCCT '15)*, pp. 169–173, IEEE, Kanyakumari, India, April 2015.

[4] B. R. Nugroho, "The architecture of an IoT-based healthcare monitoring system using smart e-health gateways in home/hospital domain," *Buletin Inovasi ICT & Ilmu Komputer*, vol. 2, no. 1, 2015.

[5] A. Agra, M. Christiansen, K. S. Ivarsøy, I. E. Solhaug, and A. Tomasgard, "Combined ship routing and inventory management in the salmon farming industry," *Annals of Operations Research*, pp. 1–25, 2016.

[6] X. Zhao, H. Fan, H. Zhu, Z. Fu, and H. Fu, "The design of the internet of things solution for food supply chain," in *Proceedings of the International Conference on Education, Management, Information and Medicine*, Shenyang, China, April 2015.

[7] S. Jabbar, M. Khan, B. Nathali Silva, and K. Han, "A REST-based industrial web of things' framework for smart warehousing," *The Journal of Supercomputing*, 2016.

[8] P. Misra, V. Rajaraman, K. Dhotrad, J. Warrior, and Y. Simmhan, "An interoperable realization of smart cities with plug and play based device management," 2015, https://arxiv.org/abs/1503.00923.

[9] M. Khan, S. Din, S. Jabbar, M. Gohar, H. Ghayvat, and S. C. Mukhopadhyay, "Context-aware low power intelligent SmartHome based on the Internet of things," *Computers & Electrical Engineering*, vol. 52, pp. 208–222, 2016.

[10] O. Aldabbas, A. Abuarqoub, M. Hammoudeh, U. Raza, and A. Bounceur, "Unmanned ground vehicle for data collection in wireless sensor networks: mobility-aware sink selection," *The Open Automation and Control Systems Journal*, vol. 8, no. 1, pp. 35–46, 2016.

[11] C. C. Grant, A. Jones, A. Hamins, and N. Bryner, "Realizing the vision of smart fire fighting," *IEEE Potentials*, vol. 34, no. 1, pp. 35–40, 2015.

[12] A. Paul, A. Ahmad, M. M. Rathore, and S. Jabbar, "Smartbuddy: defining human behaviors using big data analytics in social internet of things," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 68–74, 2016.

[13] C. Chatrapathi, M. N. Rajkumar, and V. Venkatesakumar, "VANET based integrated framework for smart accident management system," in *Proceedings of the IEEE International Conference on Soft-Computing and Network Security (ICSNS '15)*, Coimbatore, India, February 2015.

[14] A. H. Levis and L. W. Wagenhals, "C4ISR architectures: I. Developing a process for C4ISR architecture design," *Systems Engineering*, vol. 3, no. 4, pp. 225–247, 2000.

[15] D. Chen, G. Doumeingts, and F. Vernadat, "Architectures for enterprise integration and interoperability: past, present and future," *Computers in Industry*, vol. 59, no. 7, pp. 647–659, 2008.

[16] H. Park, H. Kim, H. Joo, and J. Song, "Recent advancements in the Internet-of-Things related standards: a oneM2M perspective," *ICT Express*, vol. 2, no. 3, pp. 126–129, 2016.

[17] D. Evans, "The Internet of Things: How the Next Evolution of the Internet Is Changing Everything," 2017, http://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.

[18] E. Avelar, L. Marques, D. Dos Passos, R. Macedo, K. Dias, and M. Nogueira, "Interoperability issues on heterogeneous wireless communication for smart cities," *Computer Communications*, vol. 58, pp. 4–15, 2015.

[19] M. Ganzha, M. Paprzycki, W. Pawłowski, P. Szmeja, and K. Wasielewska, "Semantic interoperability in the internet of things: an overview from the INTER-IoT perspective," *Journal of Network and Computer Applications*, vol. 81, pp. 111–124, 2017.

[20] R. Ambrosio and S. Widergren, "A framework for addressing interoperability issues," in *Proceedings of the IEEE Power Engineering Society General Meeting (PES '07)*, Tampa, Fla, USA, June 2007.

[21] K. R. Malik, T. Ahmad, M. Farhan et al., "Big-data: transformation from heterogeneous data to semantically-enriched simplified data," *Multimedia Tools and Applications*, vol. 75, no. 20, pp. 12727–12747, 2016.

[22] J. Santos, J. J. P. C. Rodrigues, B. M. C. Silva, J. Casal, K. Saleem, and V. Denisov, "An IoT-based mobile gateway for intelligent personal assistants on mobile health environments," *Journal of Network and Computer Applications*, vol. 71, pp. 194–204, 2016.

[23] A. Yachir, B. Djamaa, A. Mecheti, Y. Amirat, and M. Aissani, "A comprehensive semantic model for smart object description and request resolution in the internet of things," *Procedia Computer Science*, vol. 83, pp. 147–154, 2016.

[24] P. P. Jayaraman, D. Palmer, A. Zaslavsky, and D. Georgakopoulos, "Do-it-Yourself Digital Agriculture applications with semantically enhanced IoT platform," in *Proceedings of the 10th IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP '15)*, Singapore, April 2015.

[25] P. Desai, A. Sheth, and P. Anantharam, "Semantic gateway as a service architecture for IoT interoperability," in *Proceedings of the 3rd IEEE International Conference on Mobile Services (MS '15)*, pp. 313–319, New York, NY, USA, July 2015.

[26] A. Gyrard and M. Serrano, "Connected smart cities: interoperability with SEG 3.0 for the internet of things," in *Proceedings of the 30th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '16)*, pp. 796–802, March 2016.

[27] H. Fotouhi, A. Causevic, M. Vahabi, and M. Björkman, "Interoperability in heterogeneous low-power wireless networks for health monitoring systems," in *Proceedings of the in IEEE International Conference on Communications Workshops (ICC '06)*, 2016.

[28] E. Mingozzi, G. Tanganelli, C. Vallati, B. Martinez, I. Mendia, and M. Gonzalez-Rodriguez, "Semantic-based context modeling for quality of service support in IoT platforms," in *Proceedings of the 17th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '16)*, Coimbra, Portugal, June 2016.

[29] C. Pereira, A. Pinto, A. Aguiar, P. Rocha, F. Santiago, and J. Sousa, "IoT interoperability for actuating applications through standardised M2M communications," in *Proceedings of the 17th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '16)*, Coimbra, Portugal, June 2016.

[30] P. P. Jayaraman, C. Perera, D. Georgakopoulos, S. Dustdar, D. Thakker, and R. Ranjan, "Analytics-as-a-service in a multi-cloud environment through semantically-enabled hierarchical data processing," *Software—Practice and Experience*, 2016.

[31] N. Shariatzadeh, T. Lundholm, L. Lindberg, and G. Sivard, "Integration of digital factory with smart factory based on internet of things," *Procedia CIRP*, vol. 50, pp. 512–517, 2016.

[32] P. Buneman and S. Staworko, "RDF graph alignment with bisimulation," *Proceedings of the VLDB Endowment*, vol. 9, no. 12, pp. 1149–1160, 2016.

[33] W. Zheng, L. Zou, W. Peng, X. Yan, S. Song, and D. Zhao, "Semantic SPARQL similarity search over RDF knowledge graphs," *Proceedings of the VLDB Endowment*, vol. 9, no. 11, pp. 840–851, 2016.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 9429676, 12 pages
https://doi.org/10.1155/2017/9429676

Research Article

Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making

Bhagya Nathali Silva, Murad Khan, and Kijun Han

School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea

Correspondence should be addressed to Kijun Han; kjhan@knu.ac.kr

Received 25 October 2016; Accepted 12 December 2016; Published 18 January 2017

Academic Editor: Jaime Lloret

The concept of the smart city is widely favored, as it enhances the quality of life of urban citizens, involving multiple disciplines, that is, smart community, smart transportation, smart healthcare, smart parking, and many more. Continuous growth of the complex urban networks is significantly challenged by real-time data processing and intelligent decision-making capabilities. Therefore, in this paper, we propose a smart city framework based on Big Data analytics. The proposed framework operates on three levels: (1) data generation and acquisition level collecting heterogeneous data related to city operations, (2) data management and processing level filtering, analyzing, and storing data to make decisions and events autonomously, and (3) application level initiating execution of the events corresponding to the received decisions. In order to validate the proposed architecture, we analyze a few major types of dataset based on the proposed three-level architecture. Further, we tested authentic datasets on Hadoop ecosystem to determine the threshold and the analysis shows that the proposed architecture offers useful insights into the community development authorities to improve the existing smart city architecture.

1. Introduction

The novel concept of "connected everyday objects" over the existing network has been evolved with the emergence of the smart devices. The tremendous growth of the devices connected to the network has expanded the boundaries of conventional networks. This major breakthrough introduced Internet of Things (IoT) as the third wave of the web after static pages web (WWW) and social networking web [1, 2]. The IoT is an unceasingly growing network, capable of identifying and sharing data autonomously among heterogeneous devices, which are uniquely addressable. IoT has become the spotlight of attention among multiple interest groups due to the advancement of embedded device technology and rapid increase in the number of devices. The IoT concept has been matured with the attention of multiple interest groups and with the advancement of embedded device technology. This comes up with its productive applications like smart home, smart city, smart health, and so forth [3–6]. The smart city notion is initially coined with the aim of utilizing public services and resources efficiently to increase the quality of services offered to the urban citizens [7]. In fact, the offered services, that is, transportation, parking, surveillance, electricity, healthcare, and so forth, are optimized with the autonomous data collection via the heterogeneous devices connected to the urban IoT. It is essential to process a large amount of data on a real-time basis in order to serve the service requests efficiently. Consequent to the immense increase in data volume, the general data processing and analytical mechanisms become impotent to satisfy the real-time data processing demand. Hence, the collaboration with Big Data analytics is considered to be the ideal first step towards a smarter city. It assures flexible and real-time data processing followed by intelligent decision procedures [8]. As a result of adopting Big Data analytics to the urban IoT, this enhances the quality of services provided by the smart city.

In addition, multiple efforts have been made by academic and industrial experts to realize the notion of the

smart city. However, many efforts on individual aspects of interest are seen in the literature [9–11] covering water management, garbage management, parking management, and so forth. Therefore, complete and resilient smart city architecture has become a crucial demand, as lack of integrity deteriorates the practicability. In addition, it has to facilitate autonomous behavior, real-time data processing, real-time decision-making, and smart energy consumption and customization. Thereupon, the processing and analyzing of the colossal amount of data become a necessity. Henceforth, the urban IoT integrates Big Data analytics for the realization of the smart city [12]. For example, a smart meter at a residential building collects the meter reading that is compared with a predefined electricity consumption threshold and, based on the result, the current energy demand is notified to the smart grid. Simultaneously, consumers are notified with the current level of energy consumption, allowing them to manage the energy utilization efficiently. Indeed, the preceding scenario generates a reasonable amount of data for a single house. Moreover, data processing and decision-making should be carried out in a timely manner. Nevertheless, thousands of residential and public infrastructures in the city generate a prodigious amount of data related to a single task as mentioned above. Thus, the unification of data sources and Big Data analytics is considered to be an expedient solution to facilitate real-time operation of the smart city.

Even though the smart city has become a buzzword in the modern technological era, the actual implementation is still in its infancy. In this regard, multiple efforts are made to implement a realistic smart city. An urban IoT, "Padova Smart City," was implemented to provide ICT solutions for the city administration [7]. The framework consists of a data collection system, street lighting monitoring system, and a gateway. By means of the collected environmental parameters, that is, temperature, humidity, and light, it assures the operation of streetlights. SmartSantander test bed in North Spain is used in [8] to determine the potential benefits of Big Data analytics for smart cities. The authors have analyzed temperature, traffic, season, and working days to define a network with many interacting parts, which behave according to individual rules. Smart city architecture from a data perspective is proposed in [13]. The architecture consists of six layers covering multiple aspects of a smart city. Moreover, three-tier pyramidal architecture is proposed in [14] to facilitate transactions among heterogeneous devices across a wireless ubiquitous platform. However, most of the proposed architecture types focus on specific area of interest such as lighting, traffic congestion, and water management. Thus, the claim is valid that there is a necessity of realistic smart city architecture competent enough to make real-time intelligent decisions to uplift the quality of urban IoT services. Figure 1 presents the overview of a conventional smart city that consists of smart community, smart transportation, smart grid, smart water management, and so forth.

In this paper, Big Data analytics are integrated with the smart city architecture to propose a realistic and feasible framework for the deployment of smart cities. The proposed architecture is capable of real-time intelligent decision-making, autonomous data collecting, and user-centric energy customizing. However, the decision and control management is the most influential component for the realization of a smart city. Hence, the attainment of real-time and prompt decisions has become the utmost goal of the proposed scheme. Also, fusion techniques work to expedite the processing of the enormous amount of collected data in Big Data analytics. In this study, Hadoop is chosen as the storage and processing medium for the heterogeneous data. The Hadoop processing is followed by the generation of intelligent decisions related to the smart city operations. Finally, the actions or events corresponding to the decisions are executed.

The rest of the paper is organized as follows. Section 2 presents a detailed description of the recent literature and smart city management based on Big Data analytics. Section 3 gives a brief description of the proposed architecture. The results and analysis are presented in Section 4. Finally, the conclusion is outlined in Section 5.

2. Related Work

The rapid development of the smart city system diverts the focus of many researchers and architects towards an efficient communication and standard architectural design. Standardizing the smart city models can provide various benefits to the researchers and engineers in different contexts, naming standalone communication paradigm, detailed layering architecture, processing of information in real time, and so forth. In addition, the smart city architecture covers a variety of research approaches ranging from abstract concepts to a complete set of services. Recently, the researchers are working on deriving various solutions to present generic architecture of IoT-based smart city. Similarly, various schemes have been proposed in the current literature that follows thorough experimentation and test bed based simulations to overcome the challenges. A scheme based on experimenting a complete set of smart city services on various test bed modules has been proposed in [15]. The authors in [15] developed the physical implementation of a large-scale IoT infrastructure in a Santander city. The experimental facility is designed to be so user-friendly so that the experimenter can test the facility in different urban environments and smart city planning. A variety of new mechanisms were developed following the Santander city requirements. These mechanisms include mobility support, security and surveillance systems, large-scale support, scalability, and heterogeneity in a smart city environment. The test bed results show that the proposed architecture covers several challenges in the current literature. However, the data collected from various sensors is not tested for future urban planning and designing. Therefore, the architecture can guarantee better services in one environment but may show poor performance in another environment. Similarly, the demands of the user in an IoT-based smart environment rapidly change. Hence, it decreases the chances of understanding the context and dynamicity of the IoT-based smart user. On the other hand, the IoT is not yet matured to deploy it as generic standard for designing smart services such as smart homes and smart cities because of the following two major reasons: (1) the current IoT-based solutions are limited to specific application domain

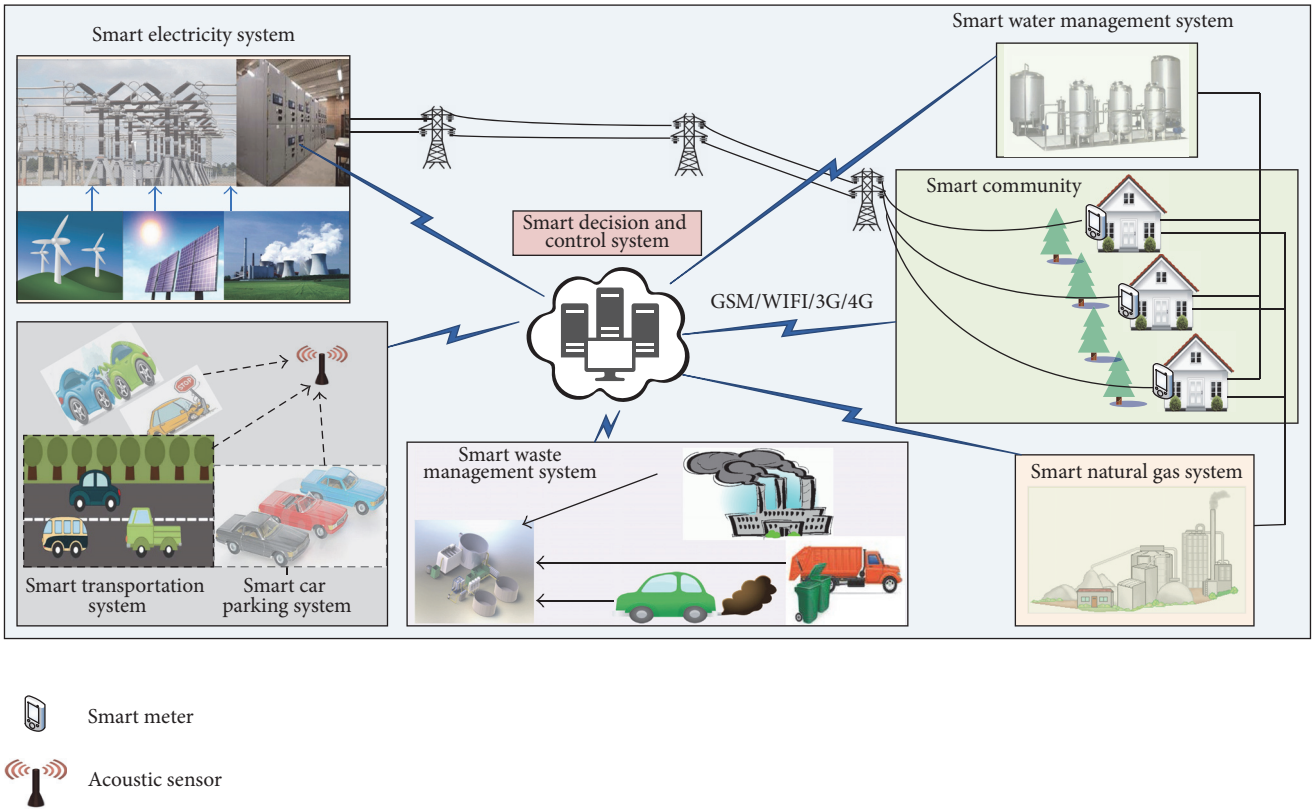


Figure 1: Typical smart city architecture.

and (2) new technologies and optimization techniques are good in one area but may be not in another. For example, wireless sensor networks (WSN) suffered high packet loss in a heterogeneous wireless environment. In addition, the deployment of IoT for one particular purpose such as waste management, air quality, noise pollution, and so forth does not reflect a standard solution [16–18]. Similarly, wireless local area network can provide low-cost services but it provides a narrow coverage compared to other technologies. Therefore, the researchers have come up with several solutions which ultimately lead to a generic communication model covering a wide set of services [19–22]. Moreover, a generic communication model can be achieved by integrating the WSN with the existing infrastructure and, thus, helps in achieving a real IoT environment with multifaceted architecture [23].

In order to design efficient and generic smart city architecture, the Big Data that is obtained from the existing smart city should be carefully examined and analyzed. The process of collection of data can be done by placing sensors in various locations in a smart home or smart city environment. Offline processing of Big Data can help in designing and planning of the urban city environment. However, it does not help in performing real-time decisions. Various techniques based on Hadoop ecosystem are developed to analyze the data for better usage and designing of the services for a smart city. For example, architecture called City Data and Analytics Platform (CiDAP) has been proposed in [8]. The authors developed layered architecture of data processing between the data sources and applications. The entire architecture consists of different parts such as data collection unit (IoT broker) and IoT agent (a repository to store data), a Big Data processing module, and a city model communication server providing the communication facilities with an external object. The data from different applications is collected and is passed to the city model server. The city model server processes the data and passes it to the IoT broker. The IoT broker separates the data based on the sensors' IDs and assigns an index number to the data. Finally, the IoT broker sends data to IoT agent for further processing. The proposed scheme achieves a higher throughput in processing of the data. Similarly, various other projects are developed based on Big Data analytics such as SCOPE [24] and FIRWARE [25]. These projects help in various aspects and provide different mechanisms to deal with Big Data in the real-time environment. However, they are not openly available to the researchers and engineers for use in different environments.

The wireless-based technologies such as wireless sensor network, wireless LAN, 3G/4G, and LTE play a vital role in providing *always best-connected* services in the smart city environment [26]. These technologies are employed in various fields and sectors of the smart city such as health care, transportation, schools, universities, and marketing. Moreover, these technologies enable a real-time communication with the smart cities devices. Thus, the data generated by the smart city sensors can be efficiently processed to take real-time decisions. However, real-time decisions require fast and efficient data processing tools. For example, Hadoop presents a solution to process the big amount of data in possible time.

Table 1: The amount of data collected in one year.

Collection frequency	1/day	1/hour	1/30 min	1/15 min
Records collected	365 m	8.75 b	17.52 b	35.04 b
Terabytes collected	1.82 tb	730 tb	1460 tb	2920 tb

m: million, b: billion, and tb: terabyte.

In addition, employing any existing tool to process Big Data depends on three properties of Big Data, that is, velocity, variety, and volume. However, processing a huge amount of data in the minimum possible time and performing real-time decision are a challenging task. Therefore, the recent research presents several models to process the data in the offline form. Thus, the outcomes can be used for management of urban planning. In order to elaborate the idea of urban planning based on Big Data analytics, we present a few example scenarios. The energy consumption recorded by smart meters in a time span of one year is shown in Table 1 [27]. The information clearly illustrates the exponential growth of data generation. The amount of data collected was calculated assuming 5 kilobytes per record [27].

The table shows that the amount of data collected by 1 million meters per 15 mins in one year is equal to 2920 TB. Thus, this high amount of data cannot be processed at once. Therefore, sophisticated tools and techniques are required to process the data and come up with proper planning and management. Similarly, processing the parking data from various parking garages in a smart city can help in designing smart parking systems. The vehicular data from various roads of a city can be used to design a smart transportation system. Moreover, this data can be used in the development of roads and bridges in various places in the smart city. Similarly, several examples of using Big Data analytics in planning and developing of smart cities services are presented in recent literature [17, 28]. However, real-time decision-making and processing on such a large amount of data are still a challenging job. In addition, an efficient smart city can be built by considering the following two points: (1) generic communication model and (2) real-time Big Data analytics.

The above literature reveals some important challenges that need to be addressed, for example, designing a generic communication model, real-time Big Data analytics, and acquisition of data from sensors in a smart city. Therefore, in this paper, we identify the need for an efficient and generic communication model for future smart cities based on Big Data analytics and integration of WSN.

3. Proposed Scheme

The proposed smart city architecture comprises three levels: (1) data generation and acquisition level, (2) data management and processing level, and (3) application level. A brief overview of the proposed smart city architecture is provided in the next subsection followed by detailed description of three levels of the proposed framework.

3.1. Overview. The layering architecture and working flow of the proposed smart city architecture are illustrated in Figure 2. Both layering and workflow are presented in a top-down manner starting from data generation and acquisition level to data management and processing level to application level. The proposed city architecture encompasses smart community development department, smart traffic control department, smart weather forecast department, and smart hospital and health department. The aforementioned components are liable for the collection of heterogeneous data within the city suburbs, thus acting as the bottom level of the proposed framework. These components are further connected with the smart decision and control system via heterogeneous access technologies such as GSM, Wi-Fi, 3G, and 4G. The autonomous decision-making uplifts the reliability as well as the practicability of the proposed scheme. Upon receiving the collected data, intelligent decisions are carried out by the smart decision and control system, situated in the middle level of the smart city framework. Moreover, the middle level regulates the events conforming to the made decisions. The event generation is taken place at the top level (application level), upon the reception of autonomous decisions.

The utmost goal of this study was to exploit realistic smart city architecture to enhance the data processing efficacy to enable real-time decision-making. In this paper, we proposed smart city architecture that incorporates Big Data analytics. In fact, there are previous studies, which integrated Big Data analytics into the smart city architecture. However, the proposed scheme is not a conventional Big Data embedded smart city as it performs explicit data filtration using Kalman filter (KF) prior to the Big Data processing. Data filtration is performed to further expedite the data processing. The KF applies threshold based filtration to distinguish between valuable and noisy data. Thus, it reduces the load that requires further processing. Similarly, we occupied a Hadoop two nodes' cluster for the Big Data processing. As shown in the Results and Data Analysis, the unification of data filtration and system architecture has enhanced the throughput of the smart city, while reducing the processing time. Thus, the proposed scheme was able to fulfill the demand for smart city architecture capable of processing data and making decision in real time.

3.2. Data Generation and Acquisition Level. A realistic smart city not only includes a prodigious amount of data but also includes complex and comprehensive computation and multiple application domains. The realization of the smart city implementation relies on all forms of data and computation due to their indispensability [13]. The smart city notion aims to optimize residential resources, to reduce traffic congestion, to provide efficient healthcare services, and to perform the water management. The acquisition of data associated with the daily operational activities become vital in terms of achieving the preceding aims. However, the data acquisition has become tedious and challenging due to the massive amount of data created by people and other connected devices. For the sake of further processing, the phenomena of interest from the real world are sensed and identified. Consequently, conversion into digital data employs various mechanisms. Low-cost and energy efficient

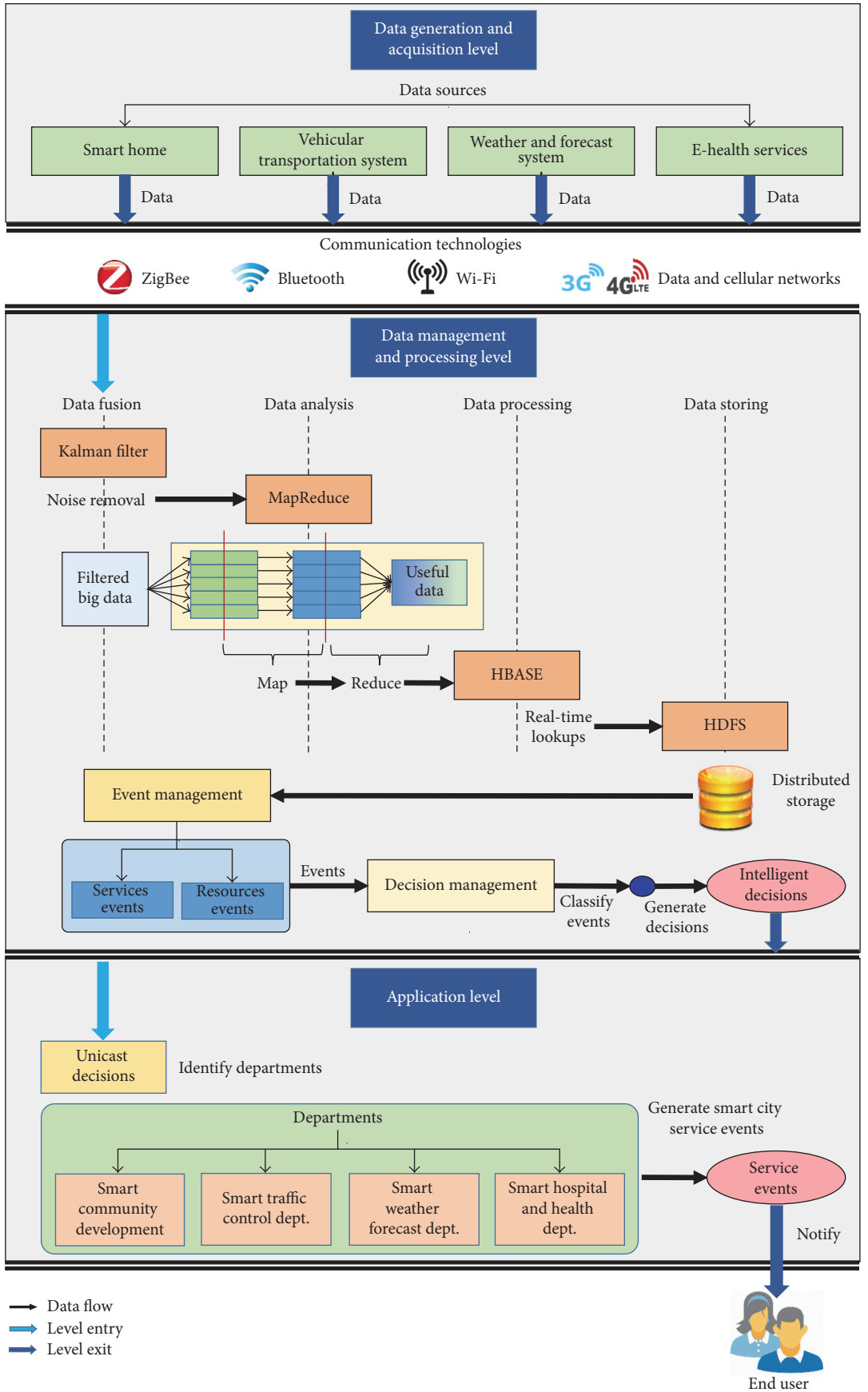


Figure 2: Working of the proposed architecture.

sensors have become a promising mechanism to acquire heterogeneous data from the urban IoT. The city becomes smarter, along with the expansion of the number of connected devices [15]. Hence, the realization of the proposed smart city architecture begins with the extensive deployment of heterogeneous sensors within the city suburbs. These sensors are liable for the collection of real-time data from the neighboring environment. The deployed context determines the type of collected data, that is, smart home, vehicular transportation system, healthcare management system, and meteorology system.

The bottom layer of the proposed scheme consists of multiple components. The key concern of the smart home is to enhance the energy utilization of the residential buildings. The home appliances are equipped with a sensor, which determines the real-time energy consumption and moves to the middle layer afterward. The data processing layer defines a threshold value for particular household's energy consumption. A data filtration process is performed by the fusion techniques to determine the values exceeding the threshold and thus optimizes further processing. Consequently, the decisions made at the middle level send to the smart community development in application level, which notifies energy consumption of a particular household to the respective residents. Meanwhile, it empowers the energy usage customization of residential buildings. The prime objective of the vehicular transportation system is to reduce the city traffic congestion. The data processing level defines the mean time that is taken to travel between two stated points. The sensors implanted on the roadsides collect vehicle entrance and departure between two points. The embedded fusion techniques determine the roads with congestion by analyzing the current travel time of stated locations, which exceeds the defined mean time. Thence, vehicular transportation system autonomously generates alternative paths and notifies the travelers via the application level. The utmost goal of the meteorology department is to ascertain the weather conditions and other environmental parameters. For example, the sensors implanted in certain locations determine the carbon monoxide (CO) concentration of the city. These sensors convey the acquired data to the middle level for filtering and processing accordingly to facilitate decision-making and event generation.

The proposed city architecture occupies multiple communication technologies; ZigBee, Bluetooth, Wi-Fi, and data and cellular networks to transmit sensed data to the data management and processing level.

3.3. Data Management and Processing Level. The data management and processing level acts as the mediator between the data acquisition and application levels. Since the crucial processes such as filtering valuable data, analyzing, processing, storing, decision-making, and generating events are carried out in this layer, this layer is considered as the brain of the proposed framework. In order to perform the aforementioned tasks, multiple modalities are embedded into this layer. Initially, the enormous amount of sensed data is filtered by fusion mechanisms to obtain valuable real-time and offline data. The MapReduce paradigm is used for the data analysis, while manipulation and storing are performed by Hadoop distributed file system (HDFS), HBASE, and HIVE.

The fusion techniques enhance the data processing efficiency by applying data filtration. Kalman filter (KF) is used to perform data filtration in the proposed framework [29]. The KF is an optimal estimator, which removes noise from the sensed data [30, 31]. The working mechanism of KF in different steps for sensor data filtration is shown as follows.

Working of KF for Sensor Data Filtration

(1) Initialization

T_k: state transition model (applied to the previous state f_{k-1})
O_k: observation model
Q_k: covariance of the process noise
R_k: covariance of the observation noise
C_k: control input model (applied to the control vector v_k)
$w_k \sim \mathcal{N}(0, Q_k)$

(2) Computing the new state f_k using the previous state f_{k-1}

$$\begin{aligned} f_k &= T_k f_{k-1} + C_k v_k + w_k \\ h_k &= O_k f_k + u_k \quad u_k \sim \mathcal{N}\left(0, R_k\right) \end{aligned} \tag{1}$$

(3) Current state estimation from the previous state

Predicted state

$$\widehat{f}_{k|k-1} = T_k \widehat{f}_{k-1|k-1} + C_k v_k \tag{2}$$

Predicted covariance

$$G_{k|k-1} = T_k G_{k-1|k-1} T_k^T + Q_k \tag{3}$$

(4) Combining current prediction with the current observation

Current observation

$$\widetilde{x}_k = h_k - O_k \widehat{f}_{k|k-1} \tag{4}$$

Observation covariance

$$S_k = O_k G_{k|k-1} O_k^T + R_k \tag{5}$$

Optimal gain

$$K_k = G_{k|k-1} O_k^T S_k^{-1} \tag{6}$$

Update state (prediction and observation)

$$\widehat{f}_{k|k} = \widehat{f}_{k|k-1} + K_k \widetilde{x}_k \tag{7}$$

(5) Update covariance (prediction and observation)

$$G_{k|k} = \left(I - K_k O_k\right) G_{k|k-1} \tag{8}$$

It initially assumes the current state f_k is evolved from the previous state f_{k-1}. The current state observation is denoted by h_k. $\widehat{f}_{k|k-1}$ represents the estimation of f at time k, while the estimation accuracy is denoted by $G_{k|k-1}$. It deduces valuable data from a large set of indirect and uncertain data. Since the KF works recursively, it processes data on arrival. Thus, it assures the real-time operation of the smart city. Moreover, it facilitates immediate processing with a minimal memory consumption. As KF removes noise from data, the data processing level utilizes its capability to infer the best estimate from a larger set of real-time data. Thereupon, the KF is manipulated to determine valuable data corresponding to the predefined threshold values. For example, the roadside sensors of the streets and roads generate a massive amount of city traffic data. Nevertheless, further processing of uncongested street data is a superfluous task. Thence, the KF determines best fitting sensed data in accordance with the predefined thresholds. Ultimately, it reduces the amount of futile data resulting in a swift analysis.

The proposed scheme stores and processes data in Hadoop framework. Thus, MapReduce has been selected as the mechanism for analyzing filtered data. MapReduce works in two steps. First is the mapping process where the set of filtered data is converted into another set of data. Next is the Reduce process which combines the data created in mapping process and results in a set of values that are reduced in amount. Data storing and processing play a major role in the realization of a smart city. As shown in Figure 2, the proposed framework utilizes multiple techniques, that is, HDFS, HBASE, HIVE, and so forth, to facilitate the above requirements. The storage demand of the proposed smart city is facilitated by HDFS, which is the primary storage of Hadoop. Since the storage of HDFS is distributed, it augments the MapReduce execution on smaller subsets of larger data cluster. In addition, HDFS assists the scalability demand of the Big Data processing. In order to favor the autonomous decision-making, the real-time read/write facility over the complete cluster is essential. Hence, HBASE is used to enhance the processing speed on Hadoop as it offers real-time lookups, in-memory caching, and server side programming. Further, it enhances the usability and the fault tolerance. HIVE provides querying and managing facility over the large amount of data that resides on the Hadoop cluster. Since SQL cannot be used to query on HIVE, we have used HiveQL to query the data on Hadoop cluster. Finally, the derived intelligent decisions are transferred to the application level of the smart city framework.

3.4. Application Level. Application level resides on top of the proposed framework. Thus, it is liable for the generation of actions corresponding to the conveyed autonomous intelligent decision. The application level is the mediator between data management level and the end user. Figure 3 presents the extended layering structure of the application level that is proposed for performance improvement of service generation. The application level is subdivided into three layers, that is, departmental layer, services layer, and subservices layer. Department layer is the boundary at the data management and processing level. Subservices layer acts as the boundary for end users. The autonomous decisions from the data processing level are unicasted to the specific departmental service, that is, smart community development department, smart traffic control department, smart weather forecast department, and smart hospital and healthcare department. The intelligent decisions of the data processing level describe the decision according to a shared vocabulary (ontology). The ontology is used to unicast the events throughout the application level. The respective departments distinguish the high-level events and the low-level events. The high-level events are stored at the departmental level and are forwarded in unicast to the recipients, whereas the low-level events are not moved further. Sequentially, the corresponding service event layer's component receives the unicast event from the departmental events. For example, the service events, smart home and waste management, are readily available to receive the departmental events from the smart community development department. Similarly, the service events are further categorized into subservice events, that is, water management and energy management under the smart home services events. The subservices events layer generates the respective event and transmits to the embedded notification component. Finally, the notification component determines the specific recipient with respect to generated event. Accordingly, it notifies the user with the generated event for the event execution.

Assume the sensors implanted on a particular city observe a street congestion. The congestion level is analyzed at the data processing level. Subsequently, the data processing level generates the appropriate intelligent decision. At the same time, the decision is communicated to the application level. The ontology determines the respective departmental event according to the decision message, that is, street congestion. Accordingly, the event is unicasted to the smart traffic control department at the application level. The departmental level determines service event component as traffic congestion. Sequentially, the generated event is forwarded to the subservice level of alternative paths. Finally, the alternative path event is notified to the respective recipient via the notification component of the application level. Moreover, the paths are notified to the potential travelers, who may enter the congested street. The smart traffic control department determines the fact by the GPS destination check and the current positioning of the vehicle.

4. Results and Data Analysis

The designing of a smart city free from existing issues entirely depends on the processing and analysis of the previous data that is obtained from various sources, that is, transportation, community department, health care, and so forth. We obtain such data from various authentic sources as is given in "*Working of KF for Sensor Data Filtration*" part. Initially, the data is fuzzy and consists of raw data entries. Therefore, on top of Hadoop system, we filter the data through KF according to our requirements that result in significant optimization of the processing time and performance efficiency of Hadoop. Moreover, the filtration process helps in processing the real-time data with less amount of time.

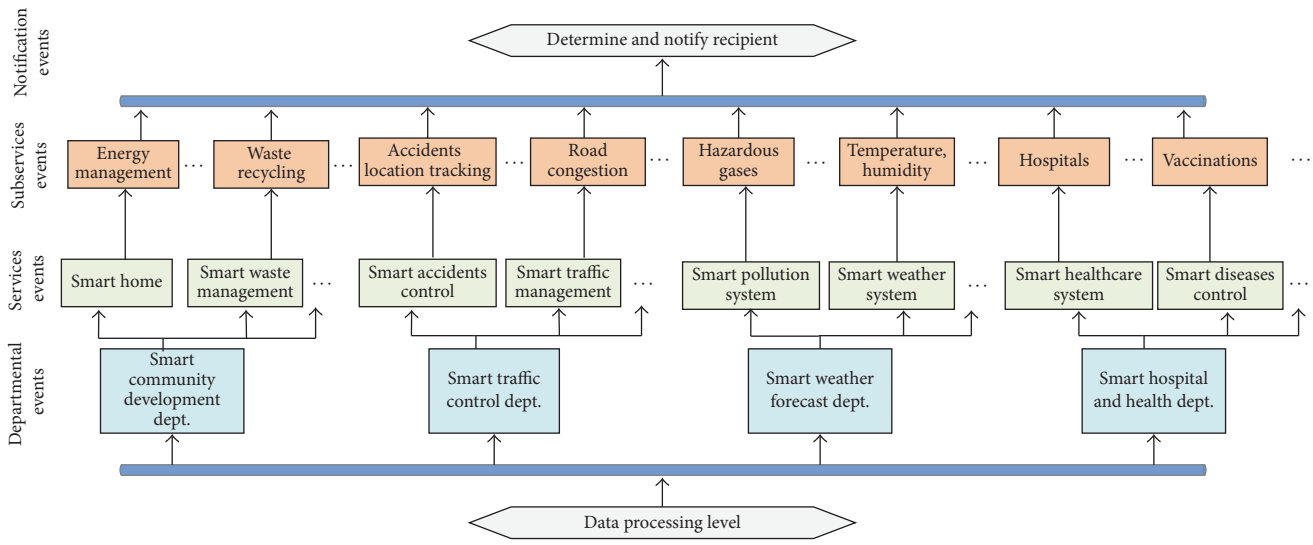


Figure 3: Event generation and processing at application level.

Table 2: Datasets information.

Sources	Dataset	Size
Surrey city, Canada [32]	Water consumption	4 MB
Aarhus city, Denmark [33]	Traffic data	3.04 GB
Aarhus city, Denmark [34]	Pollution data	77.25 MB
Aarhus city, Denmark [33]	Parking lots	0.20 MB

4.1. Dataset Information. The datasets are obtained from various authentic and reliable sources. These datasets include the following: (1) the energy and water consumption data of smart homes of survey, Canada, is obtained from the meter readings of around 61263 houses [32] and (2) the transportation and vehicular data used in the proposed analysis is obtained from number of vehicles on various roads in Aarhus city, Spain [33]. The datasets used for the analysis are openly available and authenticated. Water consumption data are covered by Open Government License of City of Surrey, Canada. Traffic data, parking lot data, and pollution data are semantically annotated datasets for the CityPulse EU FP7 project and the data is licensed under Creative Commons Attribution 4.0 International License. The dataset contains very useful information, for example, the number of vehicles and their average speed between two points on a road. (3) The parking lot dataset contain the information of various parking lots of Aarhus city, Denmark. The dataset is generated from various parking garages from May 2014 to Nov 2014 and (4) the pollution dataset consists of the information of various toxic gases such as ozone, carbon, sulfur and nitrogen dioxide, and so forth [34]. The dataset consists of the detail of some other hazardous materials but we filter out the entries of those materials for performing real-time decisions. Information corresponding to each dataset is mentioned in Table 2.

4.2. Results, Analysis, and Event Generations. The data collected from various sources is analyzed with two-node Hadoop cluster on Ubuntu 16.04 LTS having Core I5 processor and 8 GB RAM. The rationale behind this analysis is to determine normal threshold values for the actual implementation as well as to evaluate the performance of the proposed scheme in terms of processing time and data throughput. Moreover, various thresholds are defined on the output from Hadoop system. The threshold is specific to the dataset size used for the analysis. The threshold values are shown in Table 3.

Whenever the amount of data in a particular time exceeds the normal threshold, an event is generated to the respective department. In order to validate the proposed event generation system, the time taken to process the data, generate an event, and send it to the respective user (T_α) via the respective department is shown in Table 3. The experiment reveals that as the dataset size increases the time required to process it significantly increases. However, in the case of real-time processing the data is always available in streaming form and, therefore, the size of the data does not affect the system. But it is essential to make a system that directly processes the data with high speed. Therefore, the proposed filtering helps in minimizing the processing time of the Hadoop system.

The proposed system efficiently processes the data and generates various events such as traffic intensity warnings when the number of vehicles increases on a particular road compared to the normal threshold. Figure 4 illustrates number of vehicles between two defined locations for a period. The threshold value obtained from the analysis is used to determine valuable data and generate events accordingly. The threshold value was eight vehicles between two defined locations on a certain path. We set the normal threshold to 8 vehicles on a specific portion of the road. The proposed system generates the warning events in real time and sends them to the respective department. The department then broadcast the message to the vehicles headed towards that particular road. A smart parking lot system helps the citizens to get the information of empty parking lots in the surrounding

Table 3: Threshold and event generation time analysis.

Dataset	Size	Threshold	T_α
Water consumption	4 MB	80 cubic liters	11.23 s
Traffic data	3.04 GB	8 vehicles	212.88 s
Pollution data	77.25 MB	80%	16.97 s
Parking lots	0.20 MB	<10/parking garage	3.67 s

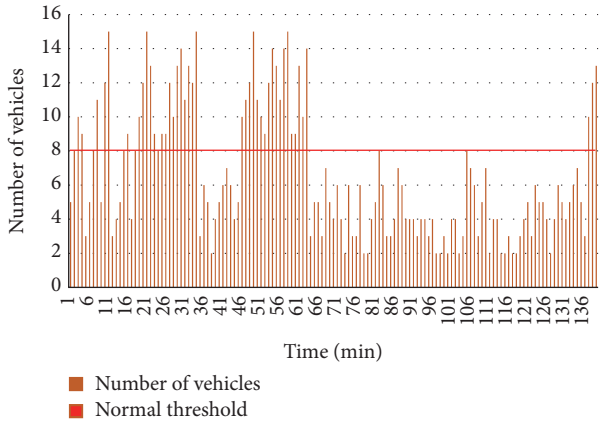


Figure 4: A number of vehicles at different roads in Aarhus, Denmark.

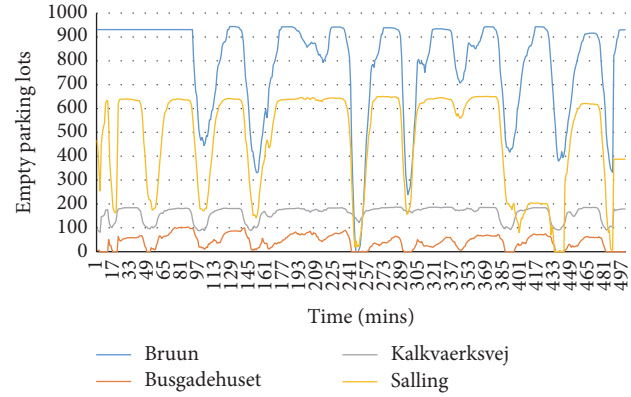


Figure 5: Empty parking lots in various places in Aarhus, Denmark.

area. The parking dataset is carefully analyzed and the empty parking lots information is sent to the respective department. The department maintains a database of empty parking lots. The citizens can check the database before parking the vehicle in the surrounding parking lots. Thus, it helps the citizens to find an empty parking lot without physical checking of the entire parking garage. Moreover, the entry for a parking lot is deleted when a citizen occupies it. Figure 5 shows various empty parking lots during a different time of the day. Initially, the Bruun city has more parking lots during daytime. However, the parking lots are rapidly filled with the passage of time. The parking lots also depend on the population and number of vehicles in a city. Thus, using the data obtain from different parking lots can be used to fulfill the parking lots requirements of the city. The shopping malls, departmental stores, and offices parking lot's data can be analyzed for better management to serve the customers with the information related to the empty parking lots. Moreover, a citizen can reserve a parking lot prior to reaching the destination.

The excessive water usage can become a critical problem in near future. Therefore, we analyzed the water usage to come up with an appropriate solution for water management. The dataset contains the water consumption information of Surrey, Canada. Figure 6 shows that each house consumed more than 80000 to 90000 liters of water each month. The normal threshold for household water consumption was obtained from the data analysis performed on the water consumption dataset. This amount of water consumption is very high and therefore in future it can become a serious problem. However, the proposed decision mechanism generates various events to the water management department to take necessary actions

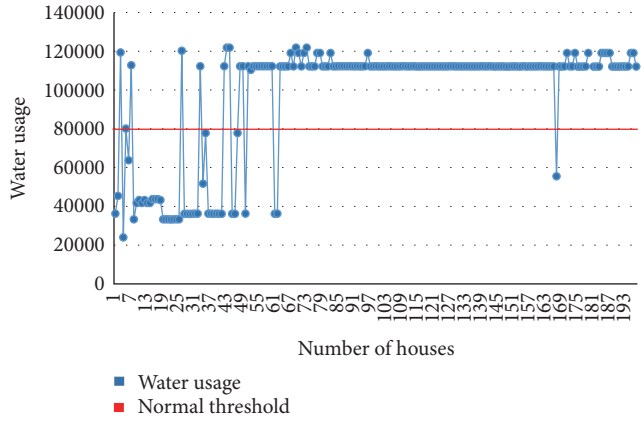


Figure 6: Water usage at different houses of Surrey, Canada.

to control the water consumption. The warning event is generated when the water consumption exceeds 80000 liters. Moreover, new water generations methods can be developed to fulfill the requirements of the citizens.

As the number of factories and vehicles is increasing, the waste production and pollution rise dramatically. Thus, controlling waste management and pollution is becoming a critical issue. In order to design a solution to handle these consequences, we analyzed the pollution data of Aarhus city at the various times of day. As shown in Figure 7, the quantity of the Ozone (O_3) is particularly high at the different times of the day. The decision system generates various events to the weather and forecast and health department to circulate a message among citizens to take great care while visiting the

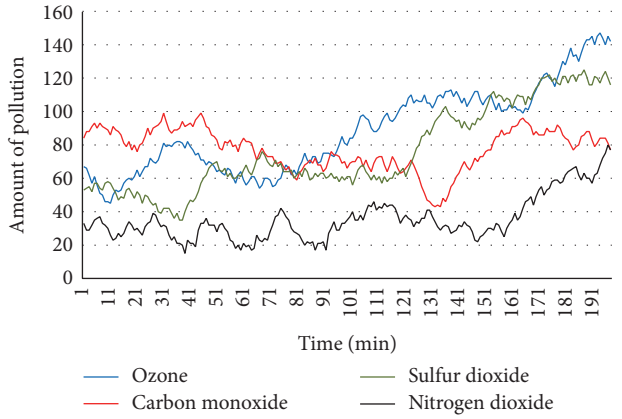


FIGURE 7: Amount of pollution at different time of day in Aarhus, Denmark.

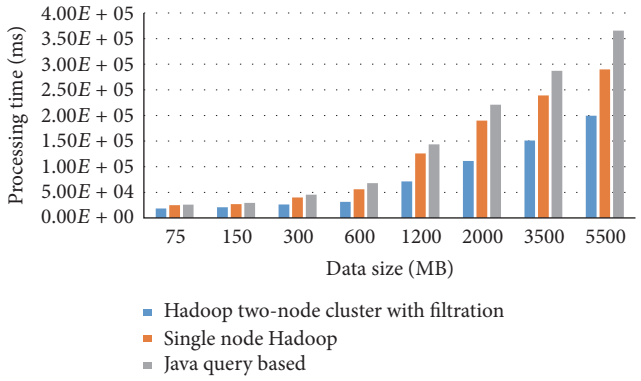


FIGURE 8: Processing time analysis.

polluted areas. Moreover, the environment control department can take necessary actions against those firms which are generating high pollution. Thus, using the pollution data can help in the future planning of the smart city as well.

The processing time of the proposed system is compared with the single node Hadoop and Java query system. The filtration of data and removing irrelevant entries from the data significantly reduce the processing time. As shown in Figure 8, as the dataset size increases, the single node Hadoop and Java query based systems required high processing time. Thus, implanting such systems in the real-time data processing environment affects the proposed decision and event generations schemes. Moreover, the efficiency of the proposed system in the context of throughput is measured as shown in Figure 9. Initially, the single Hadoop node and Java query based system process the data with similar speed. But, with the increase in dataset size, the speed of processing is highly decreased. However, the proposed scheme efficiency is significantly high compared to the single node Hadoop and Java query based system.

The analysis also shows that a city can be made smarter by analyzing data obtained from various departments. Moreover, the living style of the citizens can be improved and the comfort level can be increased by informing the citizens with

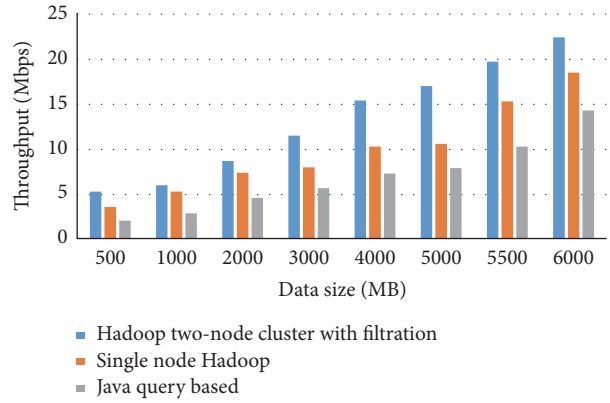


FIGURE 9: Efficiency of the system in context of throughput.

the usage of various services such as electricity and water consumption, traffic intensity on a road.

5. Conclusion and Future Work

The extensive expansion of IoT has encouraged the urban networks to be smarter, coining the notion of "smart cities." However, the realization of the smart city is still emerging, since the transformation of the conventional city operations requires novelty, networking, and processing ability of voluminous data. Therefore, the researchers and industrial experts are keen on shaping baseline architecture for a realistic smart city. In this paper, we proposed architecture of a smart city based on Big Data analytics. The key concern of this study is to ensure intelligent decision management and control center, which mediates the data acquisition sources and applications. Based on testing various types of datasets, we showed how Big Data can be used for future smart cities development and planning based on the existing data from various sources. However, our system is designed for specific goals and does not reflect a solution in general to every system present in a smart city. Moreover, a scalability option is provided in order to extend the current work in future.

As is already mentioned this work targets specific issues of a smart city to facilitate a more advanced environment for testing data in real time as well as offline. The data fusion functionality is used to reduce processing of Hadoop ecosystem on irrelevant and inappropriate data. Multiple technologies are used on top of the Hadoop storage to facilitate analysis and decision-making processes. Finally, real-world dataset of Surrey (Canada) and Aarhus (Denmark) cities are analyzed to derive the threshold values. In this study, we conceptually proposed the threefold smart city architecture for real-time decision-making. In future endeavors, we plan to carry out a simulated experiment to confirm the accuracy and efficiency of the proposed framework. Moreover, we plan to evaluate the generalizability of the proposed model, in order to standardize this smart city architecture.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This study was supported by the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering of Kyungpook National University, Korea (21A20131600005). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A1B03933566).

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: a survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for internet of things: a survey," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 70–95, 2016.

[3] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112–121, 2014.

[4] J. Pan, R. Jain, S. Paul, T. Vu, A. Saifullah, and M. Sha, "An Internet of Things framework for smart energy in buildings: designs, prototype, and experiments," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 527–537, 2015.

[5] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: a comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[6] M. Khan, B. N. Silva, and K. Han, "Internet of things based energy aware smart home control system," *IEEE Access*, vol. 4, pp. 7556–7566, 2016.

[7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[8] B. Cheng, S. Longo, F. Cirillo, M. Bauer, and E. Kovacs, "Building a big data platform for smart cities: experience and lessons from santander," in *Proceedings of the 4th IEEE International Congress on Big Data (BigData Congress '15)*, pp. 592–599, New York, NY, USA, July 2015.

[9] T.-H. Kim, R. C. Wai-Chi Fang, S. Mohammed, O. Gervasi, and A. Stoica, "Ubiquitous sensor networks and its application," *International Journal of Distributed Sensor Networks*, vol. 8, no. 7, pp. 1–3, 2012.

[10] R. E. Barone, T. Giuffrè, S. M. Siniscalchi, M. A. Morgano, and G. Tesoriere, "Architecture for parking management in smart cities," *IET Intelligent Transport Systems*, vol. 8, no. 5, pp. 445–452, 2014.

[11] M. Khan, S. Din, S. Jabbar, M. Gohar, H. Ghayvat, and S. C. Mukhopadhyay, "Context-aware low power intelligent SmartHome based on the Internet of things," *Computers & Electrical Engineering*, vol. 52, pp. 208–222, 2016.

[12] A. J. Jara, D. Genoud, and Y. Bocchi, "Big data in smart cities: from poisson to human dynamics," in *Proceedings of the 28th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '14)*, pp. 785–790, IEEE, Victoria, BC, Canada, May 2014.

[13] W. Rong, Z. Xiong, D. Cooper, C. Li, and H. Sheng, "Smart city architecture: a technology guide for implementation and design challenges," *China Communications*, vol. 11, no. 3, pp. 56–69, 2014.

[14] S. V. Nandury and B. A. Begum, "Smart WSN-based ubiquitous architecture for smart cities," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '15)*, pp. 2366–2373, Kochi, India, August 2015.

[15] L. Sanchez, L. Muñoz, J. A. Galache et al., "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, pp. 217–238, 2014.

[16] N. Maisonneuve, M. Stevens, M. E. Niessen, P. Hanappe, and L. Steels, "Citizen noise pollution monitoring," in *Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government*, Puebla, Mexico, May 2009.

[17] X. Li, W. Shu, M. Li, H.-Y. Huang, P.-E. Luo, and M.-Y. Wu, "Performance evaluation of vehicle-based mobile sensor networks for traffic monitoring," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1647–1653, 2009.

[18] T. Nuortio, J. Kytöjoki, H. Niska, and O. Bräysy, "Improved route planning and scheduling of waste collection and transport," *Expert Systems with Applications*, vol. 30, no. 2, pp. 223–232, 2006.

[19] G. Santucci, "Internet of the future and internet of things what is at stake and how are we getting prepared for them," in *Proceedings of the eMatch Conference*, pp. 1–20, Oslo, Norway, September 2009.

[20] T. Ojala, "Open urban testbed for ubiquitous computing," in *Proceedings of the International Conference on Communications and Mobile Computing (CMC '10)*, pp. 442–447, IEEE, Shenzhen, China, April 2010.

[21] S. Jabbar, A. A. Minhas, M. Imran, S. Khalid, and K. Saleem, "Energy efficient strategy for throughput improvement in wireless sensor networks," *Sensors (Switzerland)*, vol. 15, no. 2, pp. 2473–2495, 2015.

[22] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz, and P. Barnaghi, "A knowledge-based approach for real-time iot data stream annotation and processing," in *Proceedings of the 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing(CPSCom), IEEE Internet of Things (iThings)*, Taipei, Taiwan, 2014.

[23] M. Zorzi, A. Gluhak, S. Lange, and A. Bassi, "From today's INTRAnet of things to a future INTERnet of things: a wireless- and mobility-related view," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 44–51, 2010.

[24] SCOPE A Smart-city Cloud-based Open Platform and Ecosystem, https://www.bu.edu/hic/research/highlighted-sponsored-projects/scope/.

[25] FIWARE Open Source Platform, https://www.fiware.org/.

[26] M. Gohar, J.-G. Choi, S.-J. Koh, K. Naseer, and S. Jabbar, "Distributed mobility management in 6LoWPAN-based wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 620240, 12 pages, 2015.

[27] T. Dunne, *Big Data, Analytics, and Energy Consumption*, Lavastorm Agle Analytics, Boston, Mass, USA, 2012.

[28] I. J. Lee, "Big data processing framework of road traffic collision using distributed CEP," in *Proceedings of the 16th Asia-Pacific Network Operations and Management Symposium (APNOMS '14)*, Hsinchu, Taiwan, September 2014.

[29] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control Theory and Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.

[30] L. Kleeman, Understanding and Applying, http://biorobotics.ri.cmu.edu/papers/sbp_papers/integrated3/kleeman_kalman_basics.pdf.

[31] D. Li, S. Kar, J. M. F. Moura, H. V. Poor, and S. Cui, "Distributed Kalman filtering over massive data sets: analysis through large deviations of random Riccati equations," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1351–1372, 2015.

[32] Dataset Water meters, http://data.surrey.ca/dataset/water-meters.

[33] T. Dataset, Dataset Collection, http://iot.ee.surrey.ac.uk:8080/datasets.html#traffic.

[34] P. Dataset, Dataset Collection, http://iot.ee.surrey.ac.uk:8080/datasets.html#pollution.