Wireless Communications and Mobile Computing

Mobile Intelligence Assisted by Data Analytics and Cognitive Computing

Lead Guest Editor: Yin Zhang
Guest Editors: Huimin Lu and Haider Abbas

Mobile Intelligence Assisted by Data Analytics and Cognitive Computing

Wireless Communications and Mobile Computing

Mobile Intelligence Assisted by Data Analytics and Cognitive Computing

Lead Guest Editor: Yin Zhang
Guest Editors: Huimin Lu and Haider Abbas

Editorial Board

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 4302012, 2 pages
https://doi.org/10.1155/2018/4302012

Editorial

Mobile Intelligence Assisted by Data Analytics and Cognitive Computing

Yin Zhang[1], **Huimin Lu**[2] **and Haider Abbas**[3]

[1]*Zhongnan University of Economics and Law, Wuhan, China*
[2]*Kyushu Institute of Technology, Kitakyushu, Japan*
[3]*National University of Sciences and Technology, Islamabad, Pakistan*

Correspondence should be addressed to Yin Zhang; yin.zhang.cn@ieee.org

Received 17 September 2018; Accepted 17 September 2018; Published 8 November 2018

We are entering an era of big data analysis and cognitive computing. This trendy movement is observed by the pervasive use of mobile phones, storage and cloud computing, revival of artificial intelligence in practice, extended supercomputer applications, and widespread deployment of mobile devices. Although mobile devices have emerged with a great potential to change our life especially with wireless communications and mobile computing, data analytics and cognitive computing will make it possible to understand what is happening in the world more deeply.

Therefore, data analytics and cognitive computing are significant for mobile intelligence to meet many technical challenges and problems that need to be addressed to realize this potential, such as big mobile data generation and integration of multiple data sources and types. Furthermore, the mobile ecosystems have to be upgraded with new capabilities such as machine learning, data analytics, and cognitive power for providing human intelligence.

This special issue aims to explore recent advances and disseminate state-of-the-art research related to mobile intelligence on designing, building, and deploying novel technologies, to enable intelligent mobile services and applications. The response to our call for papers on this special issue was satisfactory; we finally accept 14 excellent articles covering various aspects of mobile intelligence.

In the article "Intelligent Healthcare Systems Assisted by Data Analytics and Mobile Computing," the authors present a comprehensive system design for intelligent healthcare systems assisted by data analytics and mobile computing, which consists of the data collection layer, the data management layer, and the service layer. It also introduces some representative applications based on the proposed scheme, which have been proved or demonstrated to be able to provide more intelligent, professional, and personalized healthcare services.

In the article "Performance Analysis of Location-Aware Grid-Based Hierarchical Routing Protocol for Mobile Ad Hoc Networks", the performance of a hierarchical routing protocol called Location-aware Grid-based Hierarchical Routing (LGHR) for mobile ad hoc networks (MANETs) is evaluated. In LGHR, the network is divided into nonoverlapping zones and each zone is then further divided into smaller grids.

For enhancing the femtocell capacity and mitigating both cross-tier and intratier interference in a single step, the article "Cognitive-Empowered Femtocells: An Intelligent Paradigm for Femtocell Networks" proposes a novel framework of interference management by way of channel measurement and dynamic spectrum sensing for femtocell networks, called cognitive-empowered femtocells (CEF). With the proposed framework, each CEF base station (BS) and the femtocell users can utilize spatiotemporally available radio resources for the access traffic.

The article titled "Group Recommendation Systems Based on External Social-Trust Networks" uses the trust network relationship in social networks to introduce group members' external real information, through a true evaluation of an item, to amend the group of a forecast of an item, when the group disagreement is small, that is, within the group to achieve the same case, to reduce the social network recommended to the group impact.

The article "A Multivariant Stream Analysis Approach to Detect and Mitigate DDoS Attacks in Vehicular Ad Hoc Networks" presents a novel Multivariant Stream Analysis (MVSA) approach. The proposed MVSA approach maintains the multiple stages for detection of DDoS attack in network. The approach observes the traffic in different situations and time frames and maintains different rules for various traffic classes in various time windows

The article "A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework" proposes a movie recommendation framework based on hybrid recommendation and sentiment analysis to improve the accuracy of recommender systems. The hybrid recommendation model with sentiment analysis outperforms the traditional models in terms of various evaluation criteria.

For advanced material handling, the authors of the article "Sampling Adaptive Learning Algorithm for Mobile Blind Source Separation" propose a sampling adaptive learning algorithm to calculate the adaptive learning rate in a sampling way. The algorithm has similar MSEs with adaptive step-size algorithm, but less computational time. By a smooth connection between two optimal points, the sampling method also has smooth curve and does not bring more recursion.

In the article "A Mobile Computing Method Using CNN and SR for Signature Authentication with Contour Damage and Light Distortion", the authors propose a mobile computing method of signature image authentication (SIA) with improved recognition accuracy and reduced computation time. It demonstrates theoretically and experimentally that the proposed golden global-local (G-L) algorithm has the best filtering result compared with the methods of mean filtering, medium filtering, and Gaussian filtering. The developed minimum probability threshold (MPT) algorithm produces the best segmentation result with minimum error compared with methods of maximum entropy and iterative segmentation.

The article "Framework for E-Health Systems in IoT-Based Environments" presents a specialized framework to provide smart health services in underdeveloped countries, especially in rural areas. The framework studies various aspects of IoT technology for smart health services, such as the interoperability and standardization issues, constrained and Internet environments, specialized communication protocols, and web technology requirements.

In the article "On the Tradeoff between Performance and Programmability for Software Defined WiFi Networks", the authors study software defined WiFi networks (SDWN) against traditional WiFi networks to understand the potential benefits, such as the ability of SDWN to effectively hide the handover delay between access points (AP) of the adoption of the SDWN architecture on WiFi networks and identify representative application scenarios where such SDWN approach could bring additional benefits.

Considering the on-demand MDC in cyber-physical systems, the authors of the article "On-Demand Mobile Data Collection in Cyber-Physical Systems" construct two queuing models to capture the MDC with a single MA and multiple MAs, respectively. System measures of the queues, for example, the expected values and distributions of queue length, queuing time, and response time have been explored.

The article "A Systematic Review of Security Mechanisms for Big Data in Health and New Alternatives for Hospitals" provides a view of different mechanisms and algorithms used to ensure big data security and to theoretically put forward an improvement in the health-based environment using a proposed model as reference. After analyzing the different solutions, two security alternatives are proposed combining different techniques analyzed in the state-of-the-art, with a view to providing existing information on the big data over cloud with maximum security in different hospitals located in the province of Valladolid, Spain.

The authors of the article "RADB: Random Access with Differentiated Barring for Latency-Constrained Applications in NB-IoT Network" introduce the background of NB-IoT, investigate the research on random access optimization algorithm, summarize relevant features of NB-IoT uplink and narrowband physical random access channel, and design random access with differentiated barring (RADB), which can improve the insufficiency of traditional dynamic access class barring method.

To achieve the corresponding high-level context information using the specific low-level multidomain context directly obtained by different sensors in the Internet of Things, the authors of the article "The Fusion Model of Multidomain Context Information for the Internet of Things" propose a wrapper feature selection method based on the genetic algorithm to obtain a simpler and more important subset of the context features from the low-level multidomain context, and use the decision tree algorithm which is a multiclassification algorithm to determine which high-level context the record set of the low-level context information belongs to.

We hope that this SI will serve as good reference for researches, scientists, engineers, and academicians in the field of mobile intelligence.

Conflicts of Interest

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

Acknowledgments

We would like to express our gratitude to all the authors for their generous submissions and all reviewers for their timely and professional reviews.

Yin Zhang
Huimin Lu
Haider Abbas

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 1583205, 10 pages
https://doi.org/10.1155/2018/1583205

Research Article

Performance Analysis of Location-Aware Grid-Based Hierarchical Routing Protocol for Mobile Ad Hoc Networks

Farrukh Aslam Khan[1], **Wang-Cheol Song**[2] **and Khi-Jung Ahn**[2]

[1]*Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia*
[2]*Department of Computer Engineering, Jeju National University, Jeju, Republic of Korea*

Correspondence should be addressed to Farrukh Aslam Khan; fakhan@ksu.edu.sa

Received 2 April 2018; Accepted 9 August 2018; Published 1 November 2018

Academic Editor: Yin Zhang

In this paper, the performance analysis of a hierarchical routing protocol for mobile ad hoc networks (MANETs) called Location-aware Grid-based Hierarchical Routing (LGHR) is performed. In LGHR, the network comprises nonoverlapping zones and each zone is further partitioned into smaller grids. Although LGHR is a location-aware routing protocol, the routing mechanism is similar to the link-state routing. The protocol overcomes some of the weaknesses of other existing location-based routing protocols such as Zone-based Hierarchical Link State (ZHLS) and GRID. A detailed analysis of the LGHR routing protocol is performed and its performance is compared with both the above-mentioned protocols. The comparison shows that LGHR works better than ZHLS in terms of storage overhead as well as communication overhead, whereas LGHR is more stable than GRID especially in scenarios where wireless nodes are moving with very high velocities.

1. Introduction

Mobile ad hoc networks (MANETs) are composed of wirelessly connected nodes that form a dynamic autonomous network. In MANETs, nodes communicate with one another without any centralized base station. Over the past several years, a number of MANET routing protocols have been proposed by researchers that include proactive, reactive, and hybrid routing mechanisms [1–9]. Zone Routing Protocol (ZRP) [2] is one of the hybrid routing protocols where a reactive mechanism is initiated for interzone routing and a proactive strategy is performed for intrazone routing. Another existing hybrid routing protocol called Zone-based Hierarchical Link State (ZHLS) is proposed in [4] in which all nodes in a zone communicate in a peer-to-peer manner without any central zone-head. In ZHLS, a proactive link-state routing mechanism is performed within a zone and a reactive zone-search strategy is initiated if the target node appears in a different zone. The main drawback of this routing mechanism is that each node in the network keeps the routing information of the whole zone topology, which is quite costly in case of large numbers of nodes in a zone. In the absence of a centralized authority, each node is bound to keep and update the routing table even if it is not forwarding any packets. Although ZHLS is a location-aware routing protocol, the location information kept by the GPS receiver is not utilized effectively by the protocol. For instance, if a node intends to send a packet to a destination node present in the same zone, it utilizes its intrazone routing table constructed based on the local link-state information. But if the destination lies in another zone, the source node starts a reactive zone-search strategy to obtain the destination node's zone ID by flooding packets in the whole network. In this case, if the protocol utilizes the location information obtained by the GPS receiver, it can save a number of unnecessary messages flooded throughout the network for searching the zone ID of the destination. Like other location-based routing protocols, a node can easily identify the zone ID of the destination if it knows its physical location. The location can be obtained by using a location server, as done in other location-based routing protocols such as LAR, GRID, GPSR, etc.

In GRID [10], which is a location-aware reactive routing protocol, a grid-based routing mechanism is proposed in which each grid has a gateway node and the routing is

performed only through gateway nodes in a grid-by-grid manner. The authors of GRID use the term "grid" instead of a zone. A gateway node is elected through a predefined gateway election mechanism. Like all reactive routing protocols, GRID also has a route request and a route reply mechanism to search for a route for sending packets to the destination. A major problem with this protocol is the smaller grid size. Since the main criterion for the gateway election is the shortest distance from the center of the grid, it is highly probable that the gateway nodes move out of the grid very frequently and hence, the nodes inside the grid will initiate the gateway election procedure over and over again, causing the network to become unstable. The GRID protocol does not consider any other metric for the gateway election, e.g., speed or direction of the moving gateway nodes. Moreover, since the routing is performed in a grid-by-grid manner and several grids can be there in a node's radio range, a packet will have to take several extra hops, thereby making the protocol inefficient.

Recently, we proposed a Location-aware Grid-based Hierarchical Routing (LGHR) [11] protocol for MANETs, which addressed the above-mentioned problems in ZHLS and GRID routing protocols and provided a better solution to these problems. In LGHR, each node in the network knows its own location with the help of a GPS receiver. The network is divided into nonoverlapping zones, where each zone is represented in the form of a square. Each zone is controlled by a centralized node called leader. The leader is responsible for maintaining the routing information as well as making routing decisions inside a zone. A zone is further partitioned into smaller grids, where one node is elected as a gateway node and is responsible for forwarding packets to the other nodes. The packets are routed in a gateway-by-gateway manner.

In this paper, we perform an extensive evaluation of the LGHR routing protocol. We first analyze the protocol with an example scenario, and then it is compared with both ZHLS and GRID routing protocols. For comparison with ZHLS, the numerical analysis is performed and both ZHLS and LGHR protocols are evaluated for communication as well as storage overheads. The analysis shows that LGHR performs better than ZHLS in terms of both communication and storage overheads generated by all the nodes in the network. The comparison of LGHR is also performed with GRID to analyze their stability in terms of gateway election overhead. In GRID, the election mechanism considers only the distance of a node from the center of the grid. That is, a node is elected as a gateway if it is at the shortest distance from the center of the grid. Once it is elected as a gateway, it starts functioning as a gateway until it leaves its grid. If a gateway moves out of the grid, a new election mechanism will start and another node would be elected as a gateway. In case of LGHR, not only the distance from the center of the grid is considered for electing a gateway, but the velocity of a node is also taken into consideration. This means that a node is elected as a gateway whose relative distance is minimum as compared to the other nodes. The comparison is done by performing simulations for both the protocols. The stability analysis is performed by examining the effects of several parameters on the frequency of gateway elections in a grid.

The remainder of the paper is organized as follows: Section 2 discusses the related work. The LGHR protocol is discussed in Section 3 with an example scenario, whereas Section 4 discusses the evaluation of LGHR in detail. Finally, Section 5 concludes the paper.

2. Related Work

Unlike other mobile wireless networks such as cellular and wireless IP networks having wired backbones and centralized base stations, a MANET neither has a wired backbone nor it has a centralized access point. A wireless node acts both as a host and a router. The network topology changes very frequently as the route from source to destination dynamically changes due to the node mobility. Consequently, searching for an optimal route for a destination with minimum overhead has been a challenging task for researchers over the past several years. Moreover, the limited resources in MANETs such as bandwidth and power have made the designing process of a reliable and stable routing protocol a very challenging task. A routing strategy should be able to efficiently utilize the limited resources and it should adapt to the rapidly changing network conditions.

Generally, ad hoc routing protocols can be classified into three main categories, i.e., proactive, reactive, and hybrid routing protocols. All these three kinds of routing protocols can be flat or hierarchical. Moreover, these routing protocols may also be location-aware or location-unaware. Proactive routing protocols make their routing decisions on the basis of prior topology information available, which is provided by nodes in the network. In reactive routing, a path is searched on-demand whenever there is a need to send a message to a destination. Hybrid mechanisms employ both the above strategies depending upon different criteria and situations. The architectures of these three kinds of routing protocols can be either flat or hierarchical and they can be location-aware or location-unaware. Several ad hoc routing protocols have been proposed by researchers during the past several years in the above categories. In the location-unaware category, DSDV [12], OLSR [3], and TBRPF [13] are proactive flat routing protocols, whereas STAR [14] is a proactive hierarchical routing protocol. AODV [9] and DSR [5] are reactive flat routing protocols, whereas CBRP [15] is a reactive hierarchical routing protocol. ZRP [2] is a hybrid flat routing protocol, whereas DDR [16] is a hybrid hierarchical routing protocol. In the location-aware category, DREAM [1] is classified as a proactive flat routing protocol, while LAR [7] and GPSR [6] are reactive flat routing protocols. Moreover, GRID [10] is a reactive hierarchical routing protocol, whereas ZHLS [4] is a hybrid hierarchical routing protocol. A detailed review and classification of ad hoc routing protocols can be found in [17, 18]. Some other recent MANET routing protocols can be found in [19–28].

3. Location-Aware Grid-Based Hierarchical Routing Protocol

In our location-aware hierarchical routing protocol, the leader and gateway nodes are introduced. The network is

Table 1: Neighbor table for all nodes in the example.

Neighbor Table		
Node	Position	Neighbors
1	(x1,y1)	2, 3
2	(x2,y2)	1, 3, 5, 8, 10, 19
3	(x3,y3)	1, 2
4	(x4,y4)	5, 6
5	(x5,y5)	2, 4, 6, 8
6	(x6,y6)	4, 5
7	(x7,y7)	8, 9
8	(x8,y8)	2, 5, 7, 9, 10, 19, 25, 28
9	(x9,y9)	7, 8
10	(x10,y10)	2, 8, 11, 12, 19, 25
11	(x11,y11)	10
12	(x12,y12)	10, 17, 13, 14, 19, 20,25
13	(x13,y13)	12, 14
14	(x14,y14)	13, 12
15	(x15,y15)	16, 17
16	(x16,y16)	15, 17
17	(x17,y17)	12, 15, 16, 19, 20, E
19	(x19,y19)	2, 8, 10, 12, 17, 20
20	(x20,y20)	12, 17, 19, E
21	(x21,y21)	8, 22, 25, 27
22	(x22,y22)	21, 26
23	(x23,y23)	24, 25
24	(x24,y24)	23, 25
25	(x25,y25)	8, 10, 12, 21, 23, 24, 26
26	(x26,y26)	25, 22
27	(x27,y27)	21, 28, 29, C
28	(x28,y28)	8, 27, 29
29	(x29,y29)	27, 28

partitioned into nonoverlapping zones and a central node called leader controls each zone. The responsibility of the leader is to maintain the routing information as well as make the routing decisions inside a zone. A zone is further divided into smaller grids, where an elected gateway node is responsible for forwarding packets to the other nodes. The routing is performed in a gateway-by-gateway manner. A detailed description of the working of LGHR can be found in [11].

3.1. LGHR Example Scenario. In this subsection, the proposed LGHR protocol is compared with ZHLS with the help of an example. For this purpose, consider a scenario shown in Figure 1. The leader node in LGHR constructs the neighbor table based on the information sent by the nodes inside a zone. Similarly, each zone leader sends the zone connectivity information of the neighboring zones to all the other leaders. Based on this information, the leader creates a zone table. The neighbor table for the example scenario is shown in Table 1 and the zone table is shown in Table 2(a). The neighbor table contains the neighbor node information of all the nodes. The position of each node is also written in the neighbor

Table 2: (a) Zone table for all connected zones in the network. (b) Intrazone routing table for node 8. (c) Interzone routing table maintained by node 17.

(a)

Zone Table	
Zone	Neighbor Zones
A	C, E
B	G, F
C	A, G, H
D	H, I
E	A, F
F	B, E
G	B, C
H	C, D
I	D

(b)

Intra-zone Routing Table for Node 8	
Destination	Next Node
1	2
2	2
3	2
4	5
5	5
6	5
7	8
9	8
10	10
11	10
12	10
13	10
14	10
15	19
16	19
17	19
19	19
20	19
21	21
22	21
23	25
24	25
25	25
26	21
27	28
28	28
29	28
E	19
C	28

(c)

Inter-zone Routing Table for Node 17		
Dest. Zone	Next Zone	Next Node
B	E	18
C	C	19
D	C	19
E	E	18
F	E	18
G	C	19
H	C	19
I	C	19

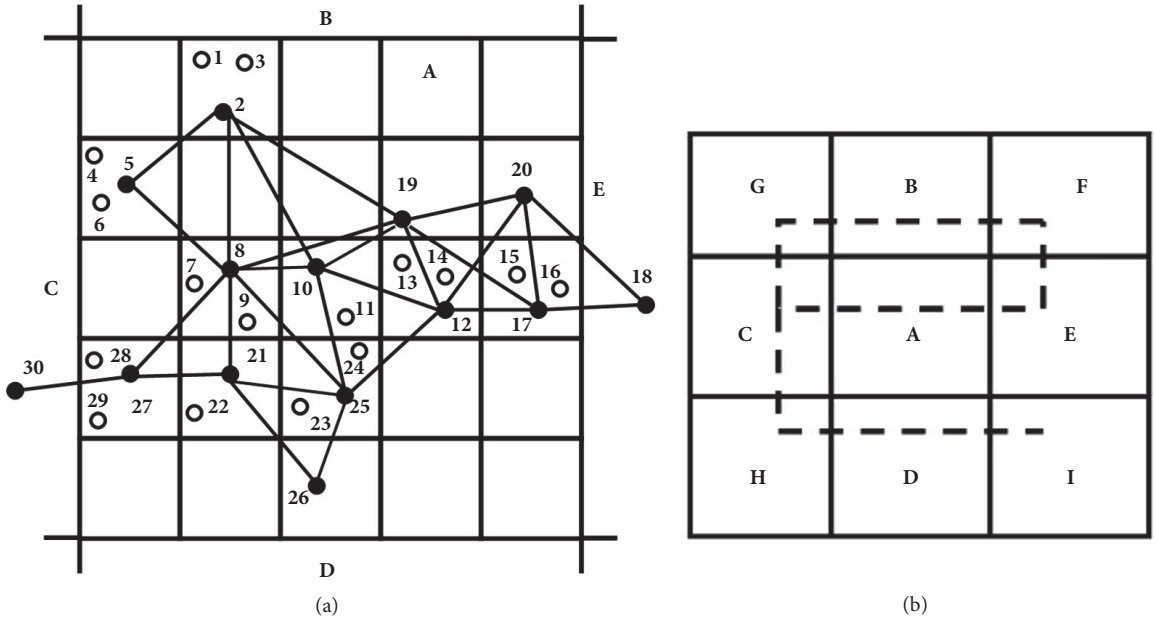


Figure 1: (a) Local topology inside zone A for our example. The connectivity of gateway nodes with other gateways is shown with solid lines. (b) Complete interzone topology for the example scenario. The dotted line shows connectivity among zones.

table, which is utilized for constructing the routing tables. In case of LGHR, any node, either gateway or nongateway, is said to be a neighbor of a nongateway node if it lies in the same grid as that of the nongateway node. Thus, the neighbor information sent by all nongateway nodes contains only the neighbor nodes that lie in their respective grids. In Table 1, it can be seen that node 1 has only two neighbors, i.e., nodes 2 and 3. Node 2 is a gateway node, whereas node 3 is a nongateway node. Secondly, neighbors of a gateway node can be nongateway nodes within its own grid and gateway nodes outside its grid that lie in its radio range. Hence, the neighbor information sent by gateway nodes contains the nongateway nodes in their respective grids as well as all the connected gateway nodes in their surrounding grids. The neighbor information sent by gateway nodes does not contain the nongateway nodes in other grids. For example, in Table 1, node 5 has nodes 2, 4, 6, and 8 as its neighbors, where node 4 and 6 are nongateway nodes within its own grid, whereas nodes 2 and 8 are gateway nodes outside its grid. Moreover, nodes 7 and 9 are not its neighbors, though they lie within its radio range. The advantage of such a criterion is to avoid extra information to be stored in the neighbor tables. Since routing is performed only in a gateway-by-gateway manner, the nongateway nodes consider only those nodes as their neighbors that lie in their respective grids. The above-mentioned criterion for a neighbor node is for situations where large numbers of nodes are present in each zone. It may be different for other situations. In case of large numbers of nodes, another possibility can be to allow only gateway nodes to send neighbor information to the leader.

The gateway nodes do not necessarily send IDs of the gateways that lie in their adjacent grids. The neighbor nodes are those gateways that come within the radio range of a gateway node. There can be a case where the neighbor node of a gateway lies in a grid not adjacent to its own grid. For example, in Figure 1(a), node 19 does not lie in the adjacent grid of node 8 but it is connected with node 8 as it lies within its radio range, and therefore, is considered to be its neighbor. The same rule applies to nodes 2 and 10. The advantage of this gateway-by-gateway routing is that although there is no node in the adjacent grid, if there is a gateway node inside the radio range of a gateway node, it can still route packets through that gateway. Moreover, if the numbers of nodes increase in the network, it will have no or very little effect on the routing performance. Since routes are computed based on the shortest path algorithm, the computed routes will always be the best routes with shortest distance in terms of the number of hops. In GRID protocol, the routing is performed in a grid-by-grid manner even if there are gateways in nonadjacent grids that lie within the radio range of a gateway node. Hence, a number of useless hops have to be taken by each packet, making the routing process inefficient.

3.1.1. Routing Table Construction. The routing table is created based on the shortest path algorithm depending upon the number of hops from the destination node. In other words, a node is selected as a next hop node, if it has the smallest number of hops from the destination node. If there is a situation where more than one path is available having same number of hops for the destination, then in that case, the physical distance is taken into consideration. Since each node that sends its neighbor information to the leader also sends its position, the physical distance between the two nodes can be easily calculated. Hence, if more than one path is available with the same number of hops, the one with the shortest physical distance from the destination would be selected. In

Table 3: Entries stored by each node and all nodes in a zone in ZHLS.

	Protocol	Node LSP Entries	Zone LSP Entries	Intra-zone Routing Table	Inter-zone Routing Table	Total Entries
Entries Stored by Each Node	ZHLS	28	9	29	8	74
Entries Stored by All Nodes	ZHLS	784	252	812	224	2072

Figure 1(a), it can be seen that if node 8 wants to send a packet to node 12, it has two paths with the same number of hops, i.e., one via node 19 and the other through node 10. In such a situation, node 8 selects node 10 as the next hop since the physical distance between node 8 and 12 is shorter using node 10 as the next hop than node 19. This can be confirmed from Table 2(b).

3.1.2. Analyzing Routing Entries. In LGHR, a leader creates and maintains both intrazone and interzone routing tables on the basis of neighbor and zone tables. Gateway nodes store their routing tables provided by the leader node but they do not keep the neighbor and zone tables. Moreover, neighbor and zone tables are created on the basis of node connectivity and zone connectivity information, which are almost similar as Node LSPs and Zone LSPs in ZHLS, respectively. Therefore, both intrazone and interzone routing tables can be computed easily for LGHR as well as for ZHLS. The local topology for the example scenario inside a zone is shown in Figure 1(a). Solid line represents the direct radio connection between two gateway nodes. Figure 1(b) shows the complete zone-level topology with all the nine zones. The dotted line tells us which zone is connected to which other zone. Considering Figure 1 as an example, the routing entries are analyzed, which are stored in Node LSPs, Zone LSPs, interzone and intrazone routing tables, in case of ZHLS and then these are compared with the entries stored by leader and gateway nodes in LGHR. For the purpose of analysis, an entry is taken as one having one row of information stored in some table in a node. The total number of bytes may differ in different entries.

In case of ZHLS, each node stores the Node LSP as well as Zone LSP and also maintains both intrazone and interzone routing tables, which is a huge burden on that node. Table 2(b) shows intrazone routing entries stored by node 8 and Table 2(c) shows interzone routing table entries stored by node 17 on the basis of the example scenario in Figure 1. The number of entries stored by these nodes as well as total entries stored by all the nodes are shown in Table 3 for ZHLS. Based on the analysis, the total numbers of entries stored by all nodes are 2072. In case of LGHR, the leader node stores neighbor and zone tables as well as intrazone and interzone routing tables of only gateway nodes as routing is performed only through gateways. *Edge Gateway* nodes store both intrazone and interzone routing tables, whereas *Intermediate Gateway* nodes store only intrazone routing tables. Table 4 shows that LGHR stores only 829 entries for one zone in this example. For the purpose of generalization, if it is assumed that the numbers of nodes are uniformly distributed in all zones and the number of gateways is also the same in all zones, then the total number of entries for 9 zones would be 18648 in case of ZHLS and 7461 in case of LGHR. This clearly shows that LGHR stores much less entries than the total entries stored by ZHLS.

4. Evaluation of LGHR

In this section, the proposed LGHR protocol is compared with two other ad hoc routing protocols, i.e., ZHLS and GRID. In case of comparison with ZHLS, the mathematical analysis is performed for both ZHLS and LGHR. Based on this analysis, the evaluation and comparison is carried out for both the protocols. The details of the mathematical analysis can be found in [11]. The comparison of LGHR with GRID cannot be fully done in all aspects, as both protocols are different in the basic routing functionality. GRID is a reactive routing protocol, whereas LGHR is a proactive routing protocol. The common thing in both the protocols is that a gateway election mechanism is carried out in each grid. It is shown that the mechanism proposed in LGHR is more robust and stable than the one shown in GRID.

4.1. Comparison with ZHLS. For evaluation, the equations of the mathematical analysis done in [11] are used. The proposed protocol LGHR is compared with ZHLS in terms of storage overhead as well as communication overhead generated.

4.1.1. Storage Overhead. Both LGHR and ZHLS protocols are compared separately for 9 and 16 gateways per zone based on the storage overhead analysis. The value for the total number of zones is in the network is 16. The numbers of nodes are increased to 1000 in the whole network. We assume that each grid in a zone contains one gateway and the gateway nodes are separated as *Edge* and *Intermediate Gateway nodes*. It can be seen that as we increase the number of nodes in the network, the number of entries stored by both the protocols also increases. However, LGHR performs better than ZHLS in all cases and stores smaller number of entries as compared to ZHLS. The results of the analysis are shown in Figure 2. These results are for one gateway in each grid. Therefore, even if the numbers of nodes are increased in LGHR, there is a very small increase in the number of entries stored, as nongateway nodes are not responsible for storing any tables. Whereas in ZHLS, each node has to store all the required entries and hence, there is a major increase in the storage overhead incurred by the protocol in case of increasing the number of nodes. In the figures, the effect on the storage overhead is shown from the

Table 4: Entries stored by leader and all gateway nodes in one zone in LGHR.

Protocol	Entries Stored By	Neighbor Table Entries	Zone Table Entries	Intra-zone Routing Table	Inter-zone Routing Table	Total Entries
LGHR	Leader	28	9	348	48	433
	All 6 Edge Gateways	0	0	174	48	222
	All 6 Intermediate Gateways	0	0	174	0	174
Total Entries stored by LGHR Protocol						829

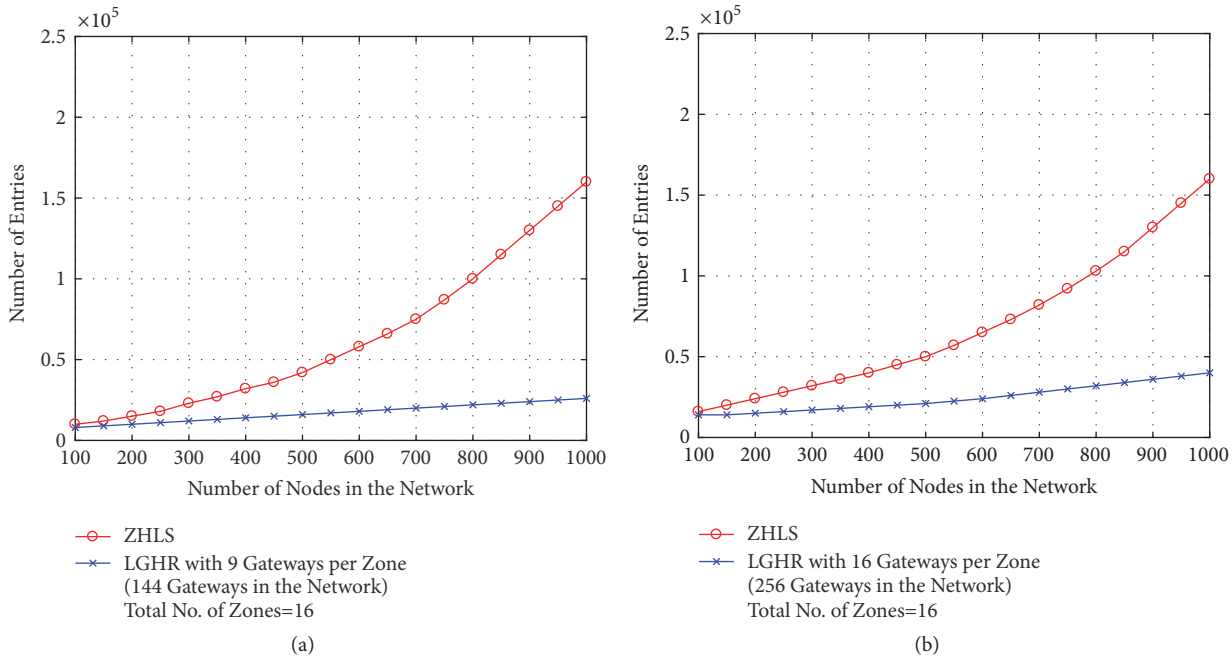


Figure 2: Comparison of LGHR with ZHLS in terms of number of entries stored for a network of 1000 nodes having 16 zones in the network. The values are shown for (a) 9 gateways per zone and (b) 16 gateways per zone.

point where the numbers of nodes are the same in both the protocols.

4.1.2. Communication Overhead. The comparison of the communication overhead for topology creation for both ZHLS and LGHR protocols is shown in Figure 3 based on the mathematical analysis. Figure 3(a) shows the comparison for 16 zones and Figure 3(b) shows the comparison for 25 zones in the network. In all cases, the communication overhead generated by LGHR is much less than that of ZHLS. The reason is that, in ZHLS, all nodes send their Node LSPs to all nodes in their zone. Similarly, each Zone LSP is sent to all the nodes. In case of LGHR, nodes in a zone are required to send their neighbor information to only the leader node. Similarly, zone tables are also propagated to only the leader nodes, not to all nodes in the network. Moreover, the leader sends the respective routing tables to only the gateway nodes. Hence, the communication overhead for topology creation by LGHR is much less than the one generated by ZHLS.

4.2. Comparison with GRID. The comparison of LGHR is performed with GRID protocol to analyze the stability of the protocol in terms of the gateway election overhead. In GRID, the election mechanism considers only the distance of a node from the center of the grid. That is, a node is elected as a gateway if it is at the shortest distance from the center of the grid. Once it is elected as a gateway, it starts functioning as a gateway until it leaves its grid. If a gateway goes out of the grid, a new election mechanism will start and another node would be elected as a gateway. In case of LGHR, not only the distance from the center of the grid is considered for electing a gateway, but the velocity of a node is also taken into consideration. This means that a node is elected as a gateway node whose relative distance from the center of the grid is minimum as compared to the other nodes. This distance is calculated by using the following equation:

$$dist_i = \sqrt{(X_i - X_c)^2 + (Y_i - Y_c)^2 + V_i^2} \quad (1)$$

Here, X_i and Y_i are the position coordinates of the ith announcing node, X_c and Y_c are the center coordinates of the grid, and V_i is the velocity of the ith node. Further details about the gateway election procedure can be found in [11]. In both the protocols, the routing is performed by gateway

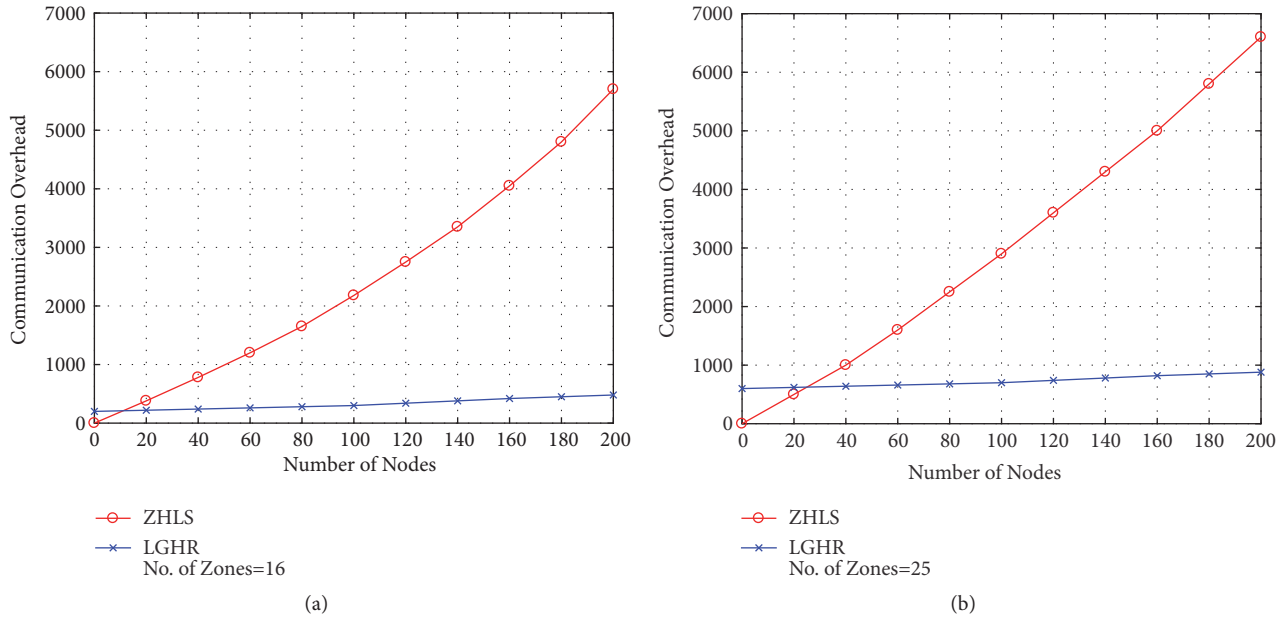


FIGURE 3: Communication overhead for topology creation generated by both LGHR and ZHLS protocols in case of (a) 16 zones in the network and (b) 25 zones in the network.

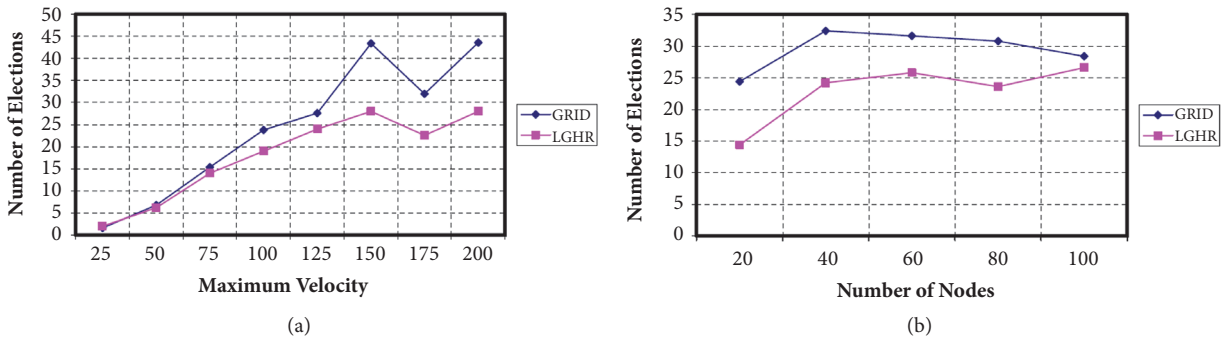


FIGURE 4: Comparison of LGHR and GRID in terms of (a) velocity of mobile nodes and (b) number of nodes in a grid.

nodes only and nongateway nodes are not responsible for forwarding packets to the other nodes; hence, the gateway node should be able to stay in the grid for longer periods of time. If the gateway node moves out of the grid quite frequently, then each time a gateway moves out, a new election mechanism will be performed. In case of mobile nodes moving with higher velocities, the gateway nodes are more likely to leave the grid frequently. Hence, the protocol will work in a more stable manner when the gateway election procedure is performed less frequently, which means that the gateway node stays inside the grid for longer periods of time. Using this criterion, the gateway election can be considered as a parameter for stability of the routing protocol.

The comparison is performed by doing simulations for both the protocols. In order to compare LGHR with GRID, the simulation environment is developed and the results are analyzed. The stability of both the protocols is analyzed by examining the effects of several parameters on the frequency of gateway elections in a grid. The parameters are, (1) velocity of nodes, (2) number of nodes in a grid, (3) size of the grid, and (4) simulation time. For all simulations, the initialization angle is taken to be 150 degrees. The curve parameter α is taken as *1*. The nodes are generated and placed in a fixed-size grid and then they are moved with given maximum velocities in random directions. Each simulation is performed five times and then the average of all the values is taken.

4.2.1. Effect of Velocity and Number of Nodes. In order to analyze the effect of velocity, the simulations are performed where the total number of nodes are 30, simulation time is 50 units, and grid size = 50 × 50. The results of keeping the number of nodes constant and increasing the velocity are shown in Figure 4(a). As shown in the figure, by increasing the velocity of mobile nodes, the number of elections for the gateway node also increases for both the protocols. This is because if nodes are moving with a higher velocity, then there is a higher probability that the nodes will go out of the grid very frequently. Hence, there will be more elections for gateway nodes for both the protocols. For the case of lower maximum velocity, both protocols perform almost the same.

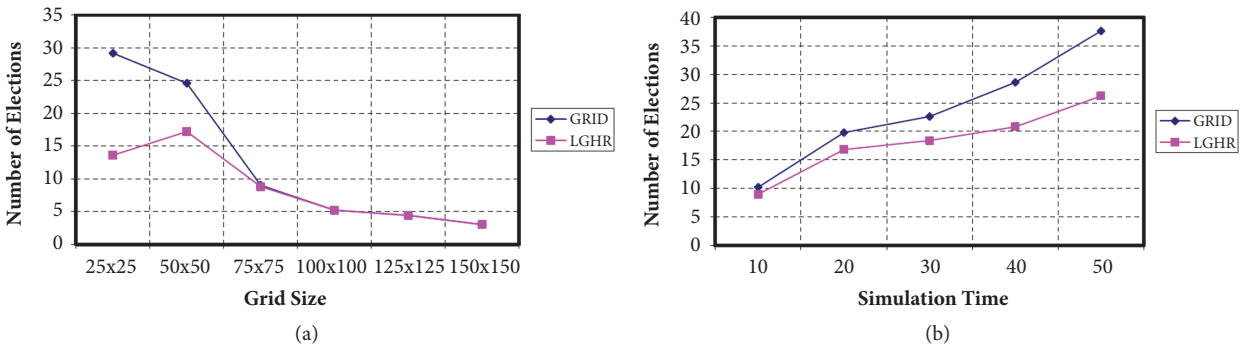


Figure 5: Comparison of LGHR and GRID in terms of (a) grid size and (b) simulation time.

As the velocity is increased, the number of elections in case of GRID starts increasing. The reason is that in GRID, there is no consideration of the velocity of mobile nodes and only the distance from the center of the grid is considered in order to elect a gateway. On the other hand, LGHR considers both the distance from the center of the grid as well as the velocity of the mobile nodes for electing a gateway node. Therefore, in case of LGHR, those nodes are elected as gateway nodes that have lower velocities and shorter distances from the center of the grid, hence making the protocol more stable.

For the second case, the following parameters are kept constant and the number of nodes is increased: maximum velocity = 150 units, grid size = 50 × 50, and simulation time = 30 units. The maximum velocity of nodes is kept constant as 150 units and the number of nodes is increased up to 100 nodes per grid. Figure 4(b) shows that by keeping the velocity constant and increasing the number of nodes, LGHR performs better than GRID. For the case when nodes are equal to 100, the difference between both the protocols becomes smaller. It is observed that if the numbers of nodes in the grid are small, then the difference between both the protocols is large. But as the numbers of nodes are increased in the grid, the difference becomes smaller between both the protocols. Since a grid is a very small part of a zone, the numbers of nodes in a grid are likely to be smaller in number. Therefore, the proposed LGHR protocol performs better than GRID in that case. The results in the figure show that even though the difference between both the protocols is small for 100 nodes, LGHR still performs better than GRID and is more stable even in that case.

4.2.2. Effect of Grid Size and Simulation Time. In order to analyze the effect of grid size in both LGHR and GRID protocols, the following parameters are kept constant: total nodes in a grid = 30, maximum velocity = 150 units, and simulation time = 20 units. Figure 5 shows that for smaller grid sizes, LGHR is more stable than GRID as it has less number of elections. As the grid size increases, the performance of both the protocols becomes similar, which means that for larger grid sizes, both the protocols work in almost the same manner. As mentioned earlier, the grid size is usually much smaller than the total size of a zone. Therefore, in realistic scenarios, for smaller grid sizes, LGHR performs better than GRID. We can see in the figure that for smaller grid sizes, more elections are likely to take place, which is due to the fact that the nodes move out of the grid very frequently. On the other hand, when the grid size is large, for example, in case of 100×100 units, the gateway elections are performed only five times in a given simulation time. Hence, the elections take place less frequently when the grid size is large.

It is observed that the duration of the simulation also affects the frequency of gateway elections. For this analysis, the following parameters are kept constant: total nodes in a grid = 30, maximum velocity = 150 units, and grid size = 50 × 50. From Figure 5(b), it is clear that the simulation time also affects the number of elections performed in a grid by both the protocols. The simulations are performed by keeping the simulation time as 10 units and then increasing it up to 50 units. It is observed that even if the simulation time is increased, LGHR still performs better than GRID. This is another indicator of the stable performance of LGHR in situations where nodes are likely to be present in the network for larger durations. The results clearly depict the effectiveness of the gateway election mechanism used in LGHR over the one used in GRID. Hence, the protocol works in a more stable manner if both the velocity and distance from the center of the grid are taken into consideration while electing the gateway node.

5. Conclusion

In this work, a detailed analysis of the Location-aware Grid-based Hierarchical Routing (LGHR) protocol is performed and the protocol is compared with two other location-aware routing protocols, i.e., ZHLS and GRID. For comparison with ZHLS, the mathematical analysis is performed and based on the analysis, both ZHLS and LGHR protocols are evaluated for storage overhead as well as for communication overhead. Moreover, the effects of increasing the number of nodes as well as zones for both the protocols are analyzed. The analysis clearly indicates that LGHR performs better than ZHLS in terms of storage overhead as well as communication overhead generated by all the nodes. ZHLS uses a hybrid approach, which may be suitable if there are small numbers of nodes in the network. However, when the nodes are increased, ZHLS incurs huge communication overhead as all nodes in a zone proactively send their link-state packets to all other nodes in that zone. Moreover, it has a reactive zone-search

mechanism, which is initiated each time a destination is found in a different zone. In LGHR, since only eligible nodes with sufficient resources can opt for becoming a leader, the burden on the leader due to carrying the routing information of other nodes can be ignored. LGHR is also compared with the GRID protocol in terms of stability. The results show that LGHR is more stable than GRID by considering the position of a node as well as its velocity for electing gateways in a grid. In all cases, LGHR outperforms GRID routing protocol and works in a more stable manner.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University, Saudi Arabia, for its funding of this research through the Research Group Project no. RGP-214.

References

[1] S. Basagni, I. Chlamtac, V. R. Syrotiuk, and B. A. Woodward, "A distance routing effect algorithm for mobility (DREAM)," in *Proceedings of the the 4th annual ACM/IEEE international conference*, pp. 76–84, Dallas, Texas, United States, October 1998.

[2] Z. J. Haas and M. R. Pearlman, "The performance of query control schemes for the zone routing protocol," in *Proceedings of the ACM SIGCOMM Conference*, vol. 28, pp. 167–177, 1998.

[3] P. Jacquet, P. Muhlethaler, and A. Qayyum, "Optimized link state routing protocol," *RFC 3626*, 2003.

[4] M. Joa-Ng and I.-T. Lu, "Peer-to-peer zone-based two-level link state routing for mobile ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1415–1425, 1999.

[5] D. B. Johnson and D. A. Maltz, *Dynamic Source Routing in Ad Hoc Wireless Networks*, Mobile Computing, Kluwer Academic Publishers, 1996.

[6] B. Karp and H. T. Kung, "GPSR: greedy Perimeter Stateless Routing for wireless networks," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM '00)*, pp. 243–254, Boston, Mass, USA, August 2000.

[7] Y.-B. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," *Wireless Networks*, vol. 6, no. 4, pp. 307–321, 2000.

[8] V. D. Park and M. S. Corson, "Temporally-ordered routing algorithm (TORA," *IETF Internet Draft*, 1999.

[9] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," No. RFC 3561, 2003.

[10] W.-H. Liao, J.-P. Sheu, and Y.-C. Tseng, "GRID: a fully location-aware routing protocol for mobile ad hoc networks," *Telecommunication Systems*, vol. 18, no. 1–3, pp. 37–60, 2001.

[11] F. A. Khan, K. Ahn, and W. Song, "A new location-aware hierarchical routing protocol for MANETs," in *Proceedings of the 7th International Conference on Ubiquitous Intelligence and Computing (UIC 2010)*, vol. 6406 of *Lecture Notes in Computer Science*, pp. 519–533, Springer Berlin Heidelberg.

[12] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proceedings of the ACM SIGCOMM'94 Conference*, vol. 24, pp. 234–244, 1994.

[13] B. Bellur, R. G. Ogier, and F. L. Templin, "Topology broadcast based on reverse-path forwarding (TBRPF)," *Internet Draft, draft-ietf-manet-tbrpf-05.txt*, 2002.

[14] J. J. Garcia-Luna-Aceves and M. Spohn, "Source-tree routing in wireless networks," in *Proceedings of the 7th International Conference on Network Protocols (ICNP '99)*, pp. 273–282, IEEE, November 1999.

[15] M. Jiang, J. Ji, and Y. C. Tay, "Cluster based routing protocol," *Internet Draft, draft-ietf-manet-cbrp-spec-01.txt*, 1999.

[16] N. Nikaein, H. Labiod, and C. Bonnet, "DDR-distributed dynamic routing algorithm for mobile ad hoc networks," in *Proceedings of the 1st Annual Workshop on Mobile and Ad Hoc Networking and Computing, MobiHOC 2000*, pp. 19–27, USA, 2000.

[17] M. Mauve, J. Widmer, and H. Hartenstein, "A survey on position-based routing in mobile ad hoc networks," *IEEE Network*, vol. 15, no. 6, pp. 30–39, 2001.

[18] M. Abolhasan, T. Wysocki, and E. Dutkiewicz, "A review of routing protocols for mobile ad hoc networks," *Ad Hoc Networks*, vol. 2, no. 1, pp. 1–22, 2004.

[19] J. Li and P. Mohapatra, "LAKER: Location aided knowledge extraction routing for mobile ad hoc networks," in *Proceedings of the 2003 IEEE Wireless Communications and Networking Conference: The Dawn of Pervasive Communication, WCNC 2003*, pp. 1180–1184, USA, March 2003.

[20] C. T. Ngo and H. Oh, "A link quality prediction metric for location based routing protocols under shadowing and fading effects in vehicular ad hoc networks," in *Proceedings of the 9th International Conference on Future Networks and Communications, FNC 2014 and the 11th International Conference on Mobile Systems and Pervasive Computing, MobiSPC 2014*, pp. 565–570, Canada, August 2014.

[21] S. Giordano and I. Stojmenovic, "Position based routing algorithms for ad hoc networks: a taxonomy," in *Ad Hoc Wireless Networking*, X. Cheng, X. Huang, and D.-Z. Du, Eds., vol. 14 of *Network Theory and Applications*, pp. 103–136, Springer, New York, NY, USA, 2002.

[22] B. Zhang and H. T. Mouftah, "Efficient grid-based routing in wireless multi-hop networks," in *Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC'05)*, Washington, DC, USA, 2005.

[23] T. Noguchi and T. Kobayashi, "Adaptive location-aware routing with directional antennas in mobile adhoc networks," in *Proceedings of the 2017 International Conference on Computing, Networking and Communications, ICNC 2017*, pp. 1006–1011, USA, January 2017.

[24] M. Kadi and I. Alkhayat, "The effect of location errors on location based routing protocols in wireless sensor networks," *Egyptian Informatics Journal*, vol. 16, no. 1, pp. 113–119, 2015.

[25] H. Aetesam and I. Snigdh, "A comparative analysis of flat, hierarchical and location-based routing in wireless sensor networks," *Wireless Personal Communications*, vol. 97, no. 4, pp. 5201–5211, 2017.

[26] T. G. Nguyen and C. So-In, "Distributed deployment algorithm for barrier coverage in mobile sensor networks," *IEEE Access*, vol. 6, pp. 21042–21052, 2018.

[27] J. N. Al-Karaki and A. Gawanmeh, "The optimal deployment, coverage, and connectivity problems in wireless sensor networks: revisited," *IEEE Access*, vol. 5, pp. 18051–18065, 2017.

[28] Sharma, M. S. Prakash, R. S. Tomar, P. K. Shrivastava, and N. Mittal, "Node connectivity and comparison between some routing protocols of MANET system," in *Recent Trends in Electronics and Communication Systems*, vol. 4, 3 edition, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3928080, 16 pages
https://doi.org/10.1155/2018/3928080

Review Article

Intelligent Healthcare Systems Assisted by Data Analytics and Mobile Computing

Xiao Ma,[1] Zie Wang,[1] Sheng Zhou,[1] Haoyu Wen,[1] and Yin Zhang [1,2]

[1]*School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, China*
[2]*State Key Laboratory for Novel Software Technology, Nanjing University, China*

Correspondence should be addressed to Yin Zhang; yinzhang@zuel.edu.cn

Received 9 January 2018; Accepted 20 May 2018; Published 3 July 2018

Academic Editor: Javier Prieto

It is entering an era of big data, which facilitated great improvement in various sectors. Particularly, assisted by wireless communications and mobile computing, mobile devices have emerged with a great potential to renovate the healthcare industry. Although the advanced techniques will make it possible to understand what is happening in our body more deeply, it is extremely difficult to handle and process the big health data anytime and anywhere. Therefore, data analytics and mobile computing are significant for the healthcare systems to meet many technical challenges and problems that need to be addressed to realize this potential. Furthermore, the advanced healthcare systems have to be upgraded with new capabilities such as machine learning, data analytics, and cognitive power for providing human with more intelligent and professional healthcare services. To explore recent advances and disseminate state-of-the-art techniques related to data analytics and mobile computing on designing, building, and deploying novel technologies, to enable intelligent healthcare services and applications, this paper presents the detailed design for developing intelligent healthcare systems assisted by data analytics and mobile computing. Moreover, some representative intelligent healthcare applications are discussed to show that data analytics and mobile computing are available to enhance the performance of the healthcare services.

1. Introduction

In the past two decades, advanced information technologies, such as mobile communication systems [1], big data [2], Internet of Things (IoT) [3], and wearable computing [4], have been widely used in the sector of healthcare [5]. Particularly, various novel healthcare systems assisted by big data and mobile computing are developed for provide intelligent and professional services [6]. However, the explosion of healthcare data brings the following challenges for data management, storage, and processing:

(i) **Large Scale**. With the improvement of electromedical and wearable devices, the data volume of healthcare systems has been extensively increasing [7].

(ii) **High Throughout**. Generally, major electromedical and wearable devices can continuously acquit health data, while these data need to be processed rapidly for prompt response to emergencies [8].

(iii) **Various Forms**. There are various healthcare data generated and stored in healthcare systems, such as medical record, hospitalization records, medical imaging, and surgery data. These multisourced data include text, image, audio, and video [9]. More importantly, the same category of healthcare data collected through different devices may follow the different data standard defined by providers.

(iv) **Deep Value**. The value of mining single source healthcare data is very limited. Thus, more research attempted to develop data fusion based approach to discover more knowledge from various data to provide more valuable services, such as personalized health guidance and public health warnings [10].

Fortunately, with the assistance of advanced techniques, more intelligent healthcare services are supported by data analytics, while it becomes more convenient for users to access to these novel services [11]. For example, M. Pramanik

et al. propose a big data enabled smart healthcare system framework to offer theoretical representations of an intra- and interorganizational business model in the healthcare context [12]. M. Rathore et al. developed a Hadoop-based intelligent healthcare system demonstrating IoT-based collaborative contextual big data sharing among all of the devices in a healthcare system [13]. S. Peddia et al. designed a cloud-based mobile e-health calorie system that can classify food objects in the plate and further compute the overall calorie of each food object with high accuracy [14].

Although the great innovation is happening in the healthcare field, there are several issues need to be addressed, especially the heterogeneous data fusion, mobile data transmission and analysis, etc. [15–17]. In [18], it discussed clear motivations and advantages of multisensor data fusion and particularly focuses on physical activity recognition, aiming at providing a systematic categorization and common comparison framework of the literature, by identifying distinctive properties and parameters affecting data fusion design choices at different levels (data, feature, and decision). In [19], it presented the electronic health record big data analytics for precision medicine, including data preprocessing, mining, and modeling.

Nowadays, a huge number of researches focus on data analysis or data mining for healthcare data [10, 20] on technical details in deploying and implementing mobile computing [21, 22], but one of the greatest challenges is how to develop a comprehensive healthcare system for effectively manage multisource heterogeneous healthcare data with particular technical features. Thus, this paper presents a detailed design of intelligent healthcare systems assisted by data analytics and mobile computing, and it make the following contributions: (1) It proposes a unified data collection layer for integrating the healthcare data from public sources and personal devices. (2) It establishes a cloud-enabled and data-driven platform for multisource heterogeneous healthcare data storage and analysis. (3) It designs a healthcare application service layer to provide unified application programming interface (API) for developers and unified interface for users.

2. Related Technologies

2.1. Mobile Computing. Mobile computing is an emerging technology related to multiple disciplines and is involved in many areas; it is also a hot issue in computer technology research. Mobile computing concerns how to provide quality information services (information storage, query, calculation, etc.) to mobile users (including users on laptops, mobile phones, and pagers) distributed across different locations. Mobile computing is a new type of technology that enables computers and other information devices to transmit data without being connected to a fixed physically connected device [23]. With the increase in mobile device usage, mobile computing is booming and has begun to be applied to different fields. Related work has been done in the education field [24].

In the medical field, mobile computing not only plays an important role but also is a very meaningful application direction as a supporting technology for mobile healthcare. Medical professionals are using mobile devices to change clinical practice. Currently, many medical software applications can help people perform nursing tasks, from information and time management to clinical decision making [25].

The use of mobile devices, healthcare professionals (HCPS), by healthcare workers has changed many aspects of clinical practice, and it has become commonplace in healthcare environments, leading to the rapid development of medical applications. Many applications can now assist healthcare professionals in many important tasks, including information and time management, health record maintenance and access, communication and counseling, information reference and gathering, patient management and monitoring, clinical decision making, medical education, and training. Current mobile devices and applications provide many benefits to HCPs, where the most significant benefit is that people can be better cared for since they support clinical decision making and help patients recover. In [25], the following benefits of mobile devices in healthcare are summarized: convenience, better clinical decision making, and improved accuracy. Mobile computing can not only improve the accuracy of identifying relevant information and the accuracy of metrics but also increase efficiency and productivity.

An effective mobile computing platform must be able to effectively use semantic-rich and medically plausible inferences about physiology, physiology, and psychology by sensing sensor information and behavioral status and linking these inferences with environmental, social, and other factors [26]. The challenge facing mobile computing is the issue of energy consumption, especially in the medical field, which often requires long monitoring durations and lengthy testing of patient-related physiological indicators. In [27], Kao et al. studied the energy consumption and performance of mobile computing, therein developing a task assignment problem and providing a dynamic programming algorithm. Hermes is a solution to the optimal strategy problem for balancing the issues of improving latency and energy consumption of mobile devices.

With the increasingly in-depth research being performed, mobile computing will have a better future in the medical field.

2.2. Big Data. Huge amounts of data are generated in the medical field, therein increasing rapidly every day, especially in mobile healthcare. Normal mobile devices and wearable devices usually need a long time to detect the related physiological indices of the human body, but mobile medicine makes obtaining a patient's physiological data more convenient and accurate, and it has greatly facilitated the development of medical big data.

Big data also provide additional data support for medical treatment, such as in medical imaging and processing, electronic health records, epidemiology, and other higher level analyses of healthcare data and can play an auxiliary role in medical diagnosis [28]. In [29], Viceconti et al. proposed that big data analysis can be successfully combined with VPH technologies to create new electronic medical solutions that are both effective and robust.

In mobile healthcare, mobile computing technologies can utilize big data technologies to perform relevant analysis and processing to obtain related data in a timely and convenient manner to provide better medical care. Lv et al. [30] introduced two mobile healthcare applications that can serve as health services based on big data. On the one hand, big data can play a role in electronic medical record collection terminals; on the other hand, big data provide doctors with solutions for developing rehabilitation tools.

In [31], the effects and benefits of big data analytics for healthcare were illustrated, therein suggesting that big data analytics may change the way healthcare professionals gain insights from their clinical and other data repositories and make informed decisions using cutting-edge technology.

However, big data technologies also face many challenges in the healthcare industry. Belle et al. [32] discussed three important up-and-coming and important areas of medical research: image-, signal-, and genomic-based analysis. In the medical field, the use of a large amount of medical data and related analytical techniques can provide a positive impact.

2.3. Cloud Computing. Cloud computing is the addition, use, and delivery of Internet-based services and often provides dynamically scalable and virtualized resources through the Internet [33]. The development of cloud computing is not limited to PCs; with the vigorous development of the mobile Internet and the emergence of various mobile terminal devices, mobile cloud computing services have emerged. Cloud computing and big data technology are inseparable and are often used in combination [34]. In [35], a networked physical system based on patient-centered healthcare applications and services was proposed and called Health CPS. It builds on cloud and big data analytics. The results of their research show that cloud and big data technologies can be used to improve the performance of medical systems and allow people to enjoy a variety of smart medical applications and services.

By combining the advantages of these two technologies, mobile cloud computing can be better applied to mobile healthcare. If certain limitations, such as limited memory, CPU power, and battery life, can be overcome, mobile cloud computing could greatly improve the capabilities and effects of mobile and cloud computing.

Loai A. et al. [36] proposed the design of networked healthcare systems using big data and mobile cloud computing technologies. Wan et al. [37] studied a cloud-enabled WBAN architecture and its applications in pervasive healthcare systems. Using energy-efficient routing, cloud resource allocation, semantic interaction, and data security mechanisms, the system transmits critical sign data to the cloud, and it provides tremendous opportunities for healthcare systems.

Certain benefits brought by cloud computing, such as the scalability offered by "potential" users (i.e., pay-as-you-go users) and software and virtual hardware services being delivered online (for example, collaborative planning, virtual servers, and virtual storage devices), will ensure that organizations do not have to maintain and update their software and hardware facilities by themselves. The flexibility of this emerging computing service can create many possibilities for organizations that currently do not exist [38].

2.4. Wearable Computing. In recent years, various mobile devices have emerged, including healthcare and medical devices (such as sports wristbands, watches, and smartphones). These wearable devices have computing abilities to record or detect relevant data for the user. Wearable computing technology has greatly facilitated mobile healthcare, as many wearable devices with disease monitoring and body perception abilities have been proposed [39, 40].

The development of wearable computing involves more aspects, and Chen et al. [41] proposed a new architecture for wearable computing based on emotional interaction and cloud technology design mechanisms; they then explored its current problems to demonstrate potential research in new directions. With the development of wearable devices, the wearable computing power needs to be further improved to better assist in medical and healthcare. Relevant studies on, e.g., computing power, connection methods, and power consumption, have been conducted [42–44]. Mobile cloud computing, big data, and other technologies are also promoting the development of wearable computing toward providing better technical support to mobile healthcare.

2.5. Internet of Things. Both wearable devices and mobile terminal devices are inseparable from the Internet of Things. The Internet of Things technology is one of the key technologies for these devices for transmitting and retrieving data and information. Its smart sensing technology provides important support for mobile medical data transmission and acquisition [45]. Catarinucci et al. [46] proposed a smart hospital system (SHS) that relies on different but complementary technologies, particularly wireless sensor networks, RFID, smartphones, and interoperability. SHS can collect real-time environmental conditions and patient physiological parameters.

Internet of Things technology, big data technology, and cloud computing are popular and widely used technologies and are often interdependent. Advanced terminal technologies (e.g., smart apparel) and advanced cloud computing technologies (e.g., big data analytics and cloud cognitive computing) are expected to provide people with more reliable and intelligent services.

In [4], Chen et al. presented a wearable medical 2.0 system to enhance QoE and QoS for the next generation of healthcare systems. In this system, washable smart clothing, including sensors, electrodes, and leads, is the key component of cloud analytics that provide users with their physiological data and receive the health and emotional state of machine-based, intelligent users. In emergency services, IoT can collect, integrate, and interoperate IoT data to flexibly support emergency medical services. The results of [47] show that the resource-based IoT data access method is effective in distributed and heterogeneous data environments and can access data across cloud and mobile computing platforms in a timely manner.

Internet of Things technology is also commonly used at home and is often connected with home healthcare [48]. The

seamless convergence of IoT devices in various platforms (such as sensors and wearable smart medicine kits) improves the user experience and service efficiency of home healthcare services (such as telemedicine).

2.6. Cyber Physical Systems. As a unity of computing processes and physical processes, cyber physical systems (CPSs) represent the next generation of intelligent systems, therein integrating computing, communications, and control [49]. The information physical system interacts with physical processes through a human-computer interaction interface and enables the networked space to manipulate a physical entity in a remote, reliable, real-time, secure, and collaborative manner.

The information physical system includes future ubiquitous environmental awareness, embedded computing, network communication, network control, and other types of system engineering; it also provides the functions of computing, communication, precise control, remote collaboration, and autonomy. In addition, such a system focuses on the close combination and coordination of computing resources and physical resources. It is mainly used in certain intelligent systems such as equipment interconnection, IoT sensors, smart homes, robotics, and intelligent navigation systems.

As a prominent subcategory of networked physical systems, to generate a convenient and cost-effective platform and promote complex and ubiquitous mobile sensing applications between humans and the surrounding physical world, mobile devices, such as smartphones, are widely used in mobile networked physical systems [50].

In [51], Costanzo et al. presented a flexible and reliable monitoring system based on embedded systems and wearable devices that allows doctors and family members to monitor the patient's distance using a cell phone. In emergency situations, proper communication with the emergency center is required to rescue patients promptly. Thus, the system can effectively monitor the health of elderly individuals.

With the rapid development of IT systems, embedded software has replaced the monitoring and diagnosis of patients. However, these systems have limitations in a wide range of device interoperability and data aggregation aspects. To solve these problems, medical network physical systems (MCP systems) have gradually become a new paradigm for medical systems.

The paradigm introduced by mobile networked physical systems (MCPs) and Time Interventions (TI) transcends traditional medical environments in terms of the monitoring, diagnosing, treating, and management of patient health. These paradigms can provide medical care to patients at any place and at any time. Among them, MCP provides the necessary technical support system that facilitates collective work in an autonomous manner.

3. Intelligent Healthcare System Architecture

3.1. Design Issues. With the increasing cost of healthcare services and medical insurance, people need more aggressive health and wellness monitoring. In the medical industry, big data and cloud computing are gradually becoming trends

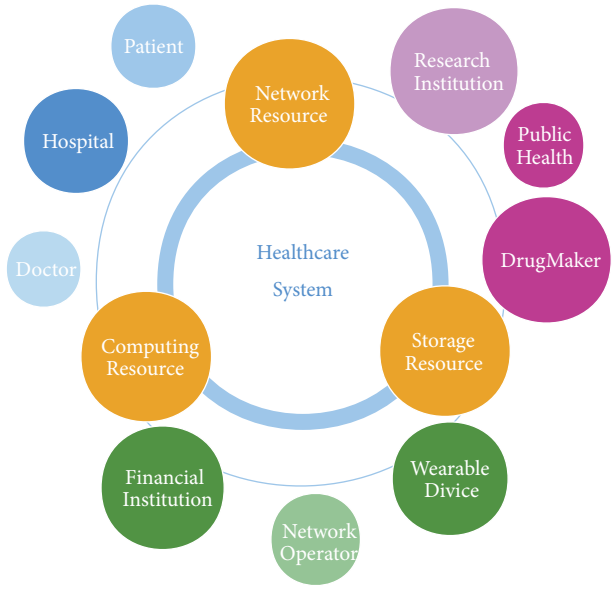


Figure 1: Expanded healthcare system.

for medical innovation. As a result, the medical industry is experiencing an increase in the amount of data generated in terms of complexity, diversity, and timeliness; the industry increasingly relies on the collection and analysis of data. Therefore, to make better decisions, we need to collect data and conduct effective analysis. The cloud is a good choice for on-demand services for storing, processing, and analyzing data. Medical data released and shared through the cloud are very popular in practice, and information and knowledge bases can be enriched and shared through the cloud.

The revolution presented by the cloud and big data can have a huge impact on the healthcare industry, and a new healthcare system is evolving. Figure 1 shows an expanded healthcare system that includes traditional roles (such as patients and healthcare providers) and other new members. This is why we need to design a more appropriate healthcare system to meet the challenges presented by this revolution.

(i) **The diversity of data sources requires a uniform standard of heterogeneous data management.** On the one hand, due to the diversification of medical equipment, the data formats and the amount of data generated by various devices may be quite different, which requires that the system support data access by various medical devices to ensure high scalability and satisfy actual medical needs. On the other hand, the system needs to convert the received data into a unified standard to improve the efficiency of data storage, query, retrieval, processing, and analysis.

(ii) **The diversity of data content requires a unified programming interface for multiple data analysis modules**. Analytical techniques have also been extended to require complex analysis for accommodating the four characteristics (4Vs) of big data in the medical field due to inconsistencies in the sources, structures, functional scenarios, and nature of health

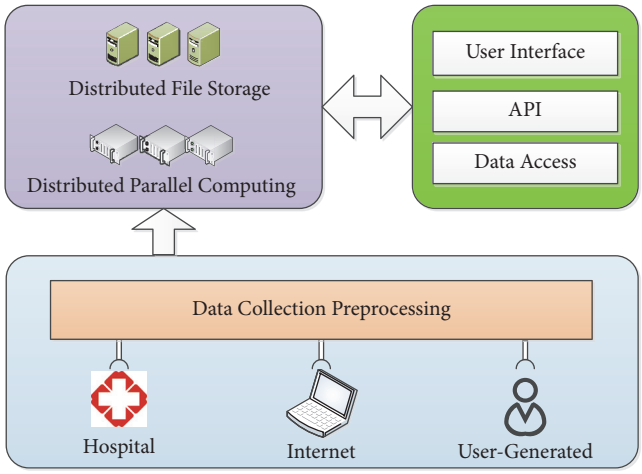


Figure 2: Architecture of intelligent healthcare systems assisted by data analytics and mobile computing.

data: volume, velocity, veracity, and variety [31]. Gone are the days of collecting data on electronic health records and other structured formats. Diversified medical data, including structured, semistructured, and other unstructured data, represent one aspect making medical data both interesting and challenging. Based on different data structures, the system can efficiently deploy and analyze data online or offline, such as via stream processing, batch processing, iterative processing, and interactive query, therein reducing system complexity and improving development and access efficiency.

(iii) **The diversity of service objects requires a uniform standard service platform interface.** Medical data that have previously been processed also have different data contents, data formats, data amounts, etc. with respect to different service targets; that is, the system can provide different applications and services with respect to the different roles of the service object. To provide reliable medical services, resource optimization, technical support, and data sharing are crucial to a system's application service platform.

3.2. Cloud- and Big-Data-Assisted Architecture. Figure 2 shows the architecture of intelligent healthcare systems assisted by data analytics and mobile computing, including the data acquisition layer, data management, and application service layer.

(i) **Data Acquisition Layer**. The main function of this layer is to collect user medical data and provide standardized data for a unified standard of multivariate data management through adapter preprocessing. This section consists of a variety of data nodes and adapters, primarily from hospitals (hospital information systems and electronic medical records), Internet (social networking services and real-time communication), and user-generated content (terminal equipment and wearable equipment). Adapters can preprocess and encrypt raw data from various devices to ensure the security and availability of data transmitted at the data management layer.

(ii) **Data Management Layer**. To support the efficient management and analysis of medical data, this layer consists of a Distributed File Storage (DFS) and Distributed Parallel Computing (DPC). DFS provides highly efficient data storage, high-throughput data upload and download, fast data retrieval. and exchange capabilities for heterogeneous medical data to improve system performance; the DPC module analyzes and processes data from the DFC module for big data. The scale of unstructured data provides offline calculations and provides the appropriate processing and analysis methods based on the timeliness of the data and the priorities of the tasks.

(iii) **Application Service Layer**. This layer consists of three parts, the user interface, API, and data access, which provides users with basic visual data analysis results. Through the data access module, following the data management layer analysis of the data, application developers can use API scheduling through the user interface to provide rich, professional, and personalized medical services.

4. Data Collection Layer

At one end, mHealth is a physician, and at the other end, it is a user with demands. The middle consists of a service provider that bridges the two ends with a variety of technologies and tools, including mobile network operators, mobile network technologies and equipment providers, mobile terminal manufacturers, IT companies (including software and hardware suppliers and system integrators), financial investors, insurance companies, public health medical institutions, banks and payment companies, private healthcare institutions, pharmaceutical companies, healthcare providers, research centers, government and nongovernmental organizations, and solution providers and more.

Data have become a particularly important aspect of mobile health. Data collection requires both the collection of devices (cell phones, computers, and portable devices) and software for gathering information. The data mainly concern the visualization of static text but can also be extended to interactive decision support algorithms, other visual image information, and communication via email and SMS integration. The consolidation of the use of geomapping components using GIS and GPS Mobile technologies is used to "tag" voice and data traffic to specific locations or to a range of locations. These combined capabilities have been used for emergency health services as well as disease surveillance, the mapping of health facilities and services, and other health-related data collection processes.

Usually, medical data classification can be divided into two aspects. One aspect is the class of data for the hospital in the operation, and the operation will produce a series of data. The other aspect is "clinical" data, which is unique to the healthcare industry. The same classification can be performed

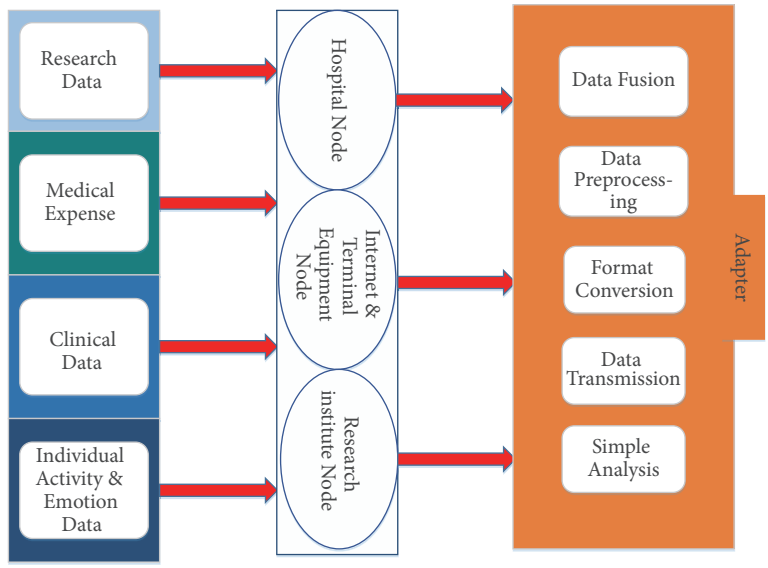


Figure 3: Data collection layer.

for personal, institutional, government, and healthcare information. The development of mobile healthcare has gradually transformed the main providers of medical data from hospitals to intermediaries or the government to make the data as public as possible without infringing upon the privacy of others. As shown in Figure 3, the data collection layer consists of data nodes and adapters.

4.1. Data Nodes. Based on mobile medical device conditions, a user-friendly intelligent healthcare system (mobile health) collects not only traditional medical data but also medical-related daily behavioral data. According to the different types of data, data nodes can be divided into the following four categories:

(i) Research organization nodes. This node mainly concerns the research and development of medical equipment companies, R & D outsourcing companies, research institutions, and other data generated during the development of the main sources of data, including (1) medical research and development processes, such as hospital clinical trials, and (2) the latest scientific research institutes. Drug R & D institutions and life science research institutions, such as major medical colleges and universities and major drug research and development centers, have accumulated a wealth of research data such as clinical trial data and high-throughput screening data. These numerical data, including personal and clinical gene and protein data, can help identify drug side effects and new effects.

(ii) Hospital nodes. In this node of the hospital, one can include most medical data, such as clinical data and medical expenses. Clinical data are a type of medical data usually collected by medical service providers for the clinical diagnosis and medical imaging of EMR, etc. These data can be unified, managed, and made open to researchers to ensure the necessary prerequisites of patient privacy and to maximize the clinical data dig value. For data generated during a hospital clinic visit, the main collection port is a medical institution such as a hospital. These data include electronic medical records, traditional test results (biochemistry, immunology, PCR, etc.), emerging test results (gene sequencing, etc.), doctors' prescriptions, and pathways for diagnosis and treatment. These data include rapidly growing segments, especially emerging test data, such as genetic test data. In addition, in this node, the medical behavior produces important cost data. The cost data refer to all audit/reimbursement records related to the payer, including patient payment records, reimbursement records, and medical circulation records such as medical expenses and medical treatment insurance claims. These data are not traditional medical data but can be used to analyze and estimate medical costs. These data are usually stored in different geographic locations in the medical institution's database using a unified data format.

(iii) Internet and endpoint devices. Personal activity and emotion as well as patient health metric data are generated at this node. These data generally refer to the patient's own, out-of-hospital behavior, and sensory data, and the main collection terminals are wearable devices and various online light medical platforms, including (1) the health management of signs collected through wearable devices data and (2) network behavior data such as registered medical consultations, online pharmacies, health management, and patient-patient exchanges. This aspect of the data can also be generated from third-party

mobile device providers (the Internet) based on mobile medical devices. These data can play a significant role in personal healthcare recommendations. Simple analysis includes personal retail consumption records reflecting lifestyle habits that can be used to assess personal health risks and develop personalized health plans. Based on the physiological data collected by wearable devices, the user's health can be easily monitored and tracked. Personal emotion data can be collected through information posted on social networks and can be used for mental health measures and emotion calculations. Particularly for the rehabilitation of patients, doctors can adjust their treatment plan based on the patient's emotions. The emotional perception of medical services with human treatments can be used to promote innovations in modern medicine [52].

4.2. Adapter. Adapter is a data node that provides access to system middleware, not simply the physical data link or the original data preprocessor and encryptor. In addition to cleaning up the data, removing redundancy, and compression, the preprocessing module also supports data format conversion. Depending on the type of data collected, the adapter uses a system-defined data standard for format conversion. The encryption module encrypts the preprocessed data to ensure security via hierarchical privacy protection. Unauthorized devices cannot decrypt packets even if they have access to the system. To improve the scalability of this system, the functional unit of the adapter is configurable. When the following basic conditions are met, the corresponding module of the adapter can be updated online.

(i) Data Node Changes. When a data node is replaced or upgraded, the functional unit will not work properly if the data format of the updated device is different from the previous version. The adapter must then send a request to the server to reconfigure the preprocessing module to be compatible with the updated module, and the server records the type of the updated data node and reauthorizes the encryption module online.

(ii) Data Standard Updates. When a new type of device without a system-defined data standard can access the system, the data standard library should be extended and is expected to be pushed to the appropriate module for updating [35].

In general, the adapter needs to implement the following five modules:

(i) Data Fusion. From the perception layer to the application layer of the Internet of Things, the types and quantities of various information are exponentially increasing. The amount of data that needs to be analyzed also increases in stages. This involves various heterogeneous networks or systems. Data fusion fuses multisensor information source data and information to obtain more accurate position estimations and identity estimations, being combined with P2P, cloud computing, and other distributed computing technologies. The above-mentioned data nodes, research data, clinical data, medical expenses, personal activity, and emotion data are effectively used to integrate, mine, and intelligently process vast amounts of data.

(ii) Preprocessing. Preprocessing refers to the preprocessing of data before the main processing. For example, before converting or enhancing most medical clinical data, the first step is to convert the irregularly distributed data into a regular distribution to facilitate operations by a computer.

(iii) Format Conversion. The data in the system change frequently. If there is no format conversion during data node replacement or upgrade, the data format will be inconsistent with the previous version; the function unit will then no longer work properly.

(iv) Data Transmission. The system is designed for user data transmission, which can be performed via WiFi, 4G, Bluetooth, and other wireless networks; for example, a user-facing system eliminates the need for patients to visit hospitals for live health data when they provide patients with personalized and intelligent monitoring of health data in real time. This can be based on cloud computing to provide intelligent health services, therein integrating a variety of wireless health monitoring equipment such as blood pressure monitors, peak flow meters, glucose monitors, scales, and ECG monitors. These devices are used by patients who need remote monitoring. The readings received from such devices are automatically forwarded to preconfigured mobile devices (mobile phones, PDAs, laptops, etc.) via Bluetooth, WiFi, and 4G cellular mobile networks. As a result, the patient's health data can be securely and accurately transmitted to a cloud-based web service without requiring the patient to have IT expertise and without user intervention [53].

(v) Simple Analysis. A simple analysis consists of the data collected from end devices, wearable devices, and the above-mentioned data nodes as well as a summary of the results. For example, some data sent by a social network, such as heartbeat condition, sleep quality, diet, fat, and calorie consumption, can be monitored by a wearable device (iWatch) to reflect the user's living conditions for implementing intelligent real-time healthcare system monitoring functionality [54].

5. Data Management Layer

As shown in Figure 4, this layer consists of a Distributed File Storage (DFS) module and a Distributed Parallel Computing (DPC) module. Using big-data-related technologies, DFS will improve system performance by providing efficient data storage and I/O for heterogeneous medical data. Based on the timeliness of the medical data and the priority of the task, DPC provides the appropriate treatment and analysis.

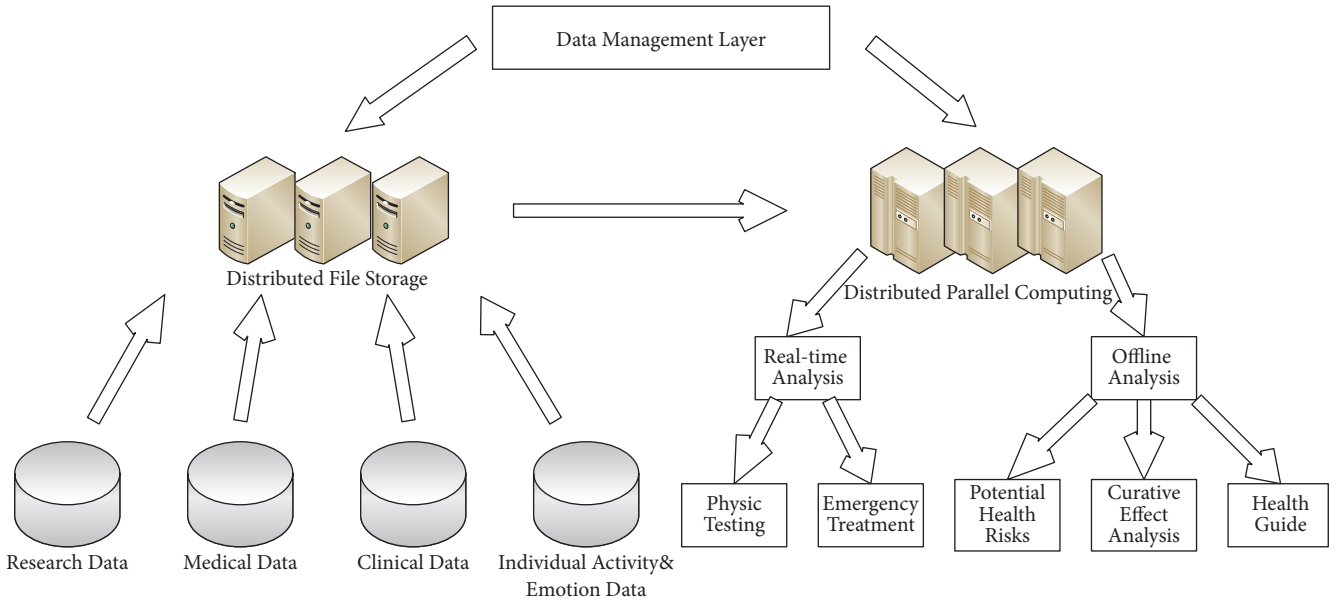


Figure 4: Data manager layer.

5.1. Distributed File Storage Module. Distributed storage systems use multiple servers to synergistically store data. Traditional network storage systems use a centralized storage server to store all the data. The storage server becomes the bottleneck for system performance and is the focus of reliability and security; traditional servers cannot meet the needs of large-scale storage applications. Distributed network storage systems adopt a scalable system architecture that utilizes multiple storage servers to share the storage load and uses the location server to locate and store information. This not only improves system reliability, availability, and access efficiency but also is easily expandable.

The distributed storage architecture consists of three parts: the client, the metadata server, and the data server. The client is responsible for sending read and write requests, cache file metadata, and file data. The metadata server is responsible for managing the metadata and processing client requests and is the core component of the entire system. The data server is responsible for storing the file data to ensure the availability and integrity and the data. The benefits of this architecture are that both performance and capacity can be expanded simultaneously, and the system is highly scalable.

Distributed storage is facing more complicated data needs, which can be divided into three categories.

Unstructured data: unstructured data include all formats of office documents, text, images, audio, and video information.

Structured data: structured data are stored in data relational libraries; one can use two-dimensional relational table structure representations. The structured data schema (schema, including attributes, data types, and the links among data) and the content is separate, and the data model needs to be predefined.

Semistructured data: between unstructured data and structured data, HTML documents belong to the semistructured data category. Such data are generally self-describing, and the biggest difference from structured data is that the schema structure and content of semistructured data are mixed, with no obvious distinction and no schema structure that predefines the data.

The main challenge facing large-scale healthcare data is how to create an efficient, mass data distributed storage mechanism and how to support efficient data processing and analysis.

For large-scale medical data, traditional relational databases obviously cannot meet big data challenges. NoSQL database [55] has a flexible model that supports easy-to-use replication, simple APIs, eventual consistency, and support for large amounts of data. This section describes the three main NoSQL databases, each based on a specific data model.

(i) Key-Value Databases. Key-value databases [55] consist of a simple data model: data key-value storage. Each key is unique, and the customer can enter the value of the query based on the key. This database structure is simple; modern key-value databases are highly scalable and have query response times that are shorter than those of relational databases. Dynamo [56] is a freely available key-value storage system, and some of Amazon's core services offer an "always-on" experience. To achieve this level of availability, Dynamo sacrifices consistency in certain failure scenarios. It extensively utilizes object-versioning and application-assisted conflict resolution in a way that offers developers a new interface.

Redis [57] is an open-source, support network, memory-based, and optionally persistent key-value pair storage database written in ANSI C. Similar to Memcached, Redis supports storing relatively many value types, including strings, lists, sets, zset (sorted set), and hashes. These data types support push pop, add/remove and intersection, union and difference sets, and richer operations, all of which are

atomic. Based on this, Redis supports a variety of different sorts. Similar to Memcached, data are cached in memory for efficiency. The difference is that Redis periodically writes updated data to the disk or writes modifications to additional log files and implements master-slave synchronization based on this.

(ii) Column-Oriented Databases. Column-oriented databases [55] store and process data based on columns rather than rows. Columnar storage allows more accurate access to the data being queried, especially in large datasets.

HBase (Hadoop Database) [58] is a high-reliability, high-performance, column-oriented, scalable distributed storage system. Just as Bigtable takes advantage of the distributed data storage provided by Google's file system, HBase provides Bigtable-like capabilities over Hadoop. HBase is a subproject of Apache's Hadoop project and is different from general relational databases, which are suitable databases for unstructured data storage.

(iii) Document Storage. Document storage [55] can support more complex data forms than key-value storage. Because documents do not follow strict patterns, pattern migration is unnecessary. In addition, key-value pairs can still be saved.

MongoDB [59] is a product between a relational database and a nonrelational database. It is the richest and most relational database among nonrelational databases. The data structure that it supports is very loose and is a json-like bson format; thus, one can store more complex data types. MongoDB's most important feature is that it supports a very powerful query language, the syntax is somewhat similar to the object-oriented query language, it can achieve most functions of a relational database single table query, and it supports the indexing of data.

Popular Distributed File Storage systems include HDFS and Colossus. The Hadoop [60] Distributed File System (HDFS) is designed to be suitable for distributed file systems running on commodity hardware. HDFS is the foundation of Hadoop applications' primary data storage, therein distributing files across 64 MB blocks of data and storing them on different nodes in a cluster for parallel computing of MR. The HDFS cluster includes a single NameNode for managing the file system's metadata and a DataNode for storing the actual data. A file is divided into one or more blocks, which are stored in the DataNode. A copy of the block is assigned to a different DataNode to prevent data loss.

Colossus [61] is the successor to the Google file system (GFS). Colossus eliminates the "single point of failure" that plagues the original Google file system. Colossus will also reduce the size of the data block to 1 MB and include multiple primary nodes, which allows Google to store more files on more machines.

5.2. Distributed Parallel Computing Module. Distributed computing is a new method of computation. The so-called distributed computing is whereby two or more software programs can share information with each other; the software can run on the same computer or on multiple computers connected to a network. This method can reduce the overall calculation time and greatly improve the computational efficiency.

The Distributed Parallel Computing module [35] analyzes and processes data from DFS and eventually discovers the information. DPC provides offline computing for a large amount of unstructured data, supports real-time data analysis and querying, and integrates various data mining and machine learning algorithms. DPC supports both real-time analysis and offline analysis.

Real-time analysis [35]: in the context of intensive care, emergency disease detection, or vital sign monitoring, changing data reflect personal health statuses in real time. Therefore, these data need to be promptly processed, and the results of the analysis are expected to be quickly responded to in case of emergencies. Through a memory analysis framework, heart rate, blood pressure, and other related data as well as other important data are recorded to improve the efficiency of the analysis.

Offline analysis [35]: in situations where there are no strong response time requirements (e.g., health assessments, medical recommendations, and health planning), common offline analysis methods in DPC, including machine learning, statistical analysis and recommendation algorithms, can be used to provide individuals with a potential health risk assessment, efficacy analysis, health guidance, etc.

Mainstream distributed computing frameworks include Hadoop, Apache Spark, BOINC, and Apache Storm.

The Hadoop framework [62] transparently provides reliability and data movement for applications. Hadoop implements a programming paradigm called MapReduce: applications are partitioned into many smaller parts, and each part can be run or rerun on any node in the cluster. In addition, Hadoop provides a distributed file system to store data for all compute nodes, bringing very high bandwidths to the entire cluster. MapReduce and the distributed file system design allow the entire framework to automatically handle node failures. It connects applications with thousands of individually calculated computers and petabytes of data.

Apache Spark [63] is an open-source cluster computing framework originally developed by the University of California, Berkeley, AMPLab. MapReduce, in comparison to Hadoop, stores the mediation data to disk after it has finished working. Spark uses in-memory arithmetic to analyze the data in memory when the data have not been written to the hard disk. Spark runs programs in memory at up to 100x faster than Hadoop MapReduce. Spark can even run up to 10x faster when running programs on the hard disk. Spark allows users to load data into the cluster memory and query them many times, making Spark ideal for machine learning algorithms.

BOINC (Berkeley open network computing platform) [64] is a mainstream distributed computing platform and is a distributed computing system developed by the University of California at Berkeley Computer Science Department. Originally designed for the SETI@home project, the fields currently using BOINC include mathematics, medicine, astronomy, and meteorology. BOINC brings together computers and mobile devices from volunteers around the world to provide computing power to researchers.

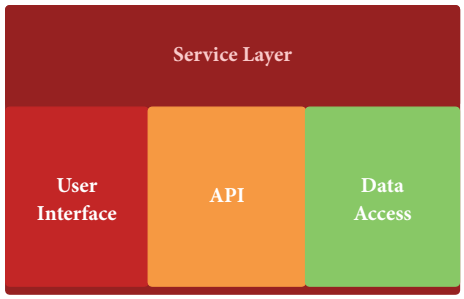


Figure 5: The function of service layer.

Apache Storm [65] is a distributed computing framework written mainly in the Clojure programming language. Originally created by Nathan Marz and his team at BackType, the project opened up after being made available on Twitter. It uses user-created "spouts" and "bolts" to define sources and operations to enable the streaming of data in bulk and in a distributed manner. The Storm application is designed to be a topology whose interface creates a transformation "flow". It provides similar functionality as a MapReduce job, and in the event of an exception, the topology theoretically runs indefinitely until it is manually terminated.

6. Service Layer

6.1. Main Function of Service Layer. The system is expected to provide a variety of applications and services for different roles (hospitals, patients, wearable device manufacturers, research institutes, pharmaceutical manufacturers, etc.)

As shown Figure 5, the main function of the service layer has the following three points:

(i) **User Interface**. The user interface provides a unified interface for the user, which makes the interaction and exchange of information between the user and system convenient and provides rich, professional, and personalized medical services.

(ii) **API**. The API provides a unified application programming interface for developers, which makes programming easy for developers.

(iii) **Data Access**. Medical data not only come from multiple sources, such as hospitals, research institutes, pharmaceutical companies, and patients, but also have different structures, i.e., being structured, semistructured, or unstructured. Data access provides a unified data access interface for these multisource heterogeneous data.

6.2. Framework of Service Layer. As shown in Figure 6, the service layer framework consists of three parts, namely, the operating platform, the management platform, and the development platform.

The operating platform is the foundation of the service layer, providing the essential resources for running healthcare applications, i.e., hardware, software and data. Hardware can include memory, software can include application software and operating systems, and data can include personal health data, clinical test data, and data on the efficacy of medicine.

The management platform is responsible for managing various applications in the system, including configuration management, deployment management, optimization management, monitoring management, visualization management, and privilege management.

(i) Configuration management is responsible for managing configuration parameters related to the system such as configuration parameter changes.

(ii) Deployment management is responsible for deploying environments and components, which are necessary for system operation.

(iii) Optimization management is responsible for configuring various types of resources within the system efficiently and selecting the final combination that most improves system performance.

(iv) Monitoring management is responsible for monitoring system operations in real time, monitoring user requests, and making judgments on the priority of requests to ensure the stable and normal operation of the system.

(v) Visualization management is responsible for providing multiple modes and ways of displaying data because there may be different targets or target users, and developers need a wide range of graphical tools.

(vi) Privilege management is responsible for assigning authority in a system that provides services for a variety of roles, and different roles have different permissions.

The development platform is responsible for providing a unified API, data access, and a testing platform for developers. Unified API: to reduce the complexity of the system, improve efficiency, and allow developers to more easily program, providing a unified API is necessary.

Data access: ddata are an essential factor in an application. Data access is a process during which developers are allowed to link to the data source through the application to access the data and process the data after returning to the application.

Testing platform: the testing platform is a platform for developers to evaluate the quality of application products and detect weaknesses in the application.

6.3. Data-Oriented Healthcare Services. According to their technical complexity and commercial value, the applications can be divided into the following four groups:

(i) **Statistics-based application services** only provide basic statistics and report services. The general approach is to first determine a time period, form the statistics of the data within the time period, and finally draw the corresponding report. Personal health status reports are representative applications. For example, it is well known that there is a health app on the Apple iPhone. Whether you are calculating carbohydrates, calories, caffeine, or other important

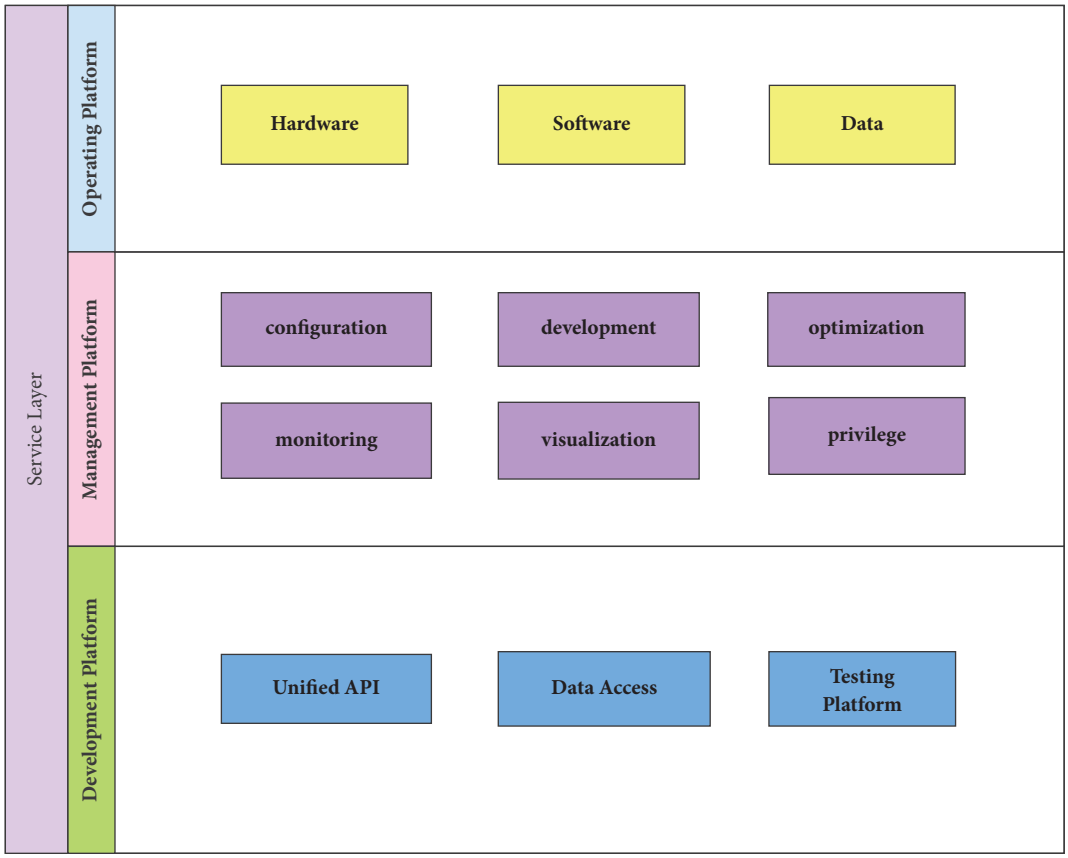


Figure 6: The framework of the service layer.

nutritional indicators, health apps can easily manage your goals, check your daily dietary intake, and graphically display them. Therefore, you can keep track of whether your nutritional intake is up to standards and understand the advantages and disadvantages of your diet. In addition, the statistical information posted on social networks reflects the emotional status of individuals and can be used in mental health measures and emotional calculations. For example, De Choudhury M. et al. [66] presented a statistical methodology to identify whether an individual engaged in mental health discourse on social media is likely to transition to that around suicidal ideation in the future. They leveraged a large dataset from a number of mental health and suicide support communities on Reddit to address their research problem. Their method can detect the obvious signs of turning to a suicidal ideation.

(ii) **Monitoring-based application services** are often used to monitor individual vital signs. Through real-time analysis, one can discover a user's physiological changes in time to avoid sudden illnesses. For example, Luca Catarinucci et al. [46] proposed a novel, IoT-aware, intelligent system for automatically monitoring and tracking patients and personnel and biomedical devices within hospitals and care facilities. The system can monitor a patient's physiological parameters and environmental conditions in real time and send them to the control center. The control center analyzes the data received and sends an alert message after an exception is detected. Through offline analysis of historical data, recovery procedures can be tracked to support treatment optimization. For example, Giancarlo Fortino et al. [67] proposed a system architecture, Body Cloud, which integrates Body Sensor Network (BSN) services with the cloud computing infrastructure. Body Cloud is an SaaS architecture that supports the storage and management of sensor data streams and offline analysis of stored data using software services hosted in the cloud, thereby allowing physicians to make timely adjustments for treatment plans.

(iii) **Knowledge-based application services** are the most representative of big data applications. With the support of data mining and machine learning technologies, data dependencies and dependencies can be found. Typical applications include the diagnosis of chronic diseases, genetic disease analysis, treatment evaluation, side effect identification, and public health alerts. For genetic disease analysis, Bravo et al. [68] proposed a novel text mining system called Be Free, which aims to identify the relationships between

biomedical entities, with particular attention paid to genes and their associated diseases. By exploiting the morphosyntactic information of the text, Be Free is able to identify gene-disease, drug-disease, and drug-target associations with state-of-the-art performance. They demonstrate the value of the gene-disease associations extracted by Be Free through a number of analyses and integration with other data sources. Be Free succeeds in identifying genes associated with a major cause of morbidity worldwide, depression, which are not present in other public resources. In [69], Lu et al. propose a unified big data processing framework based on the level set evolution method for wound image segmentation, to maximize the advantages of traditional level set models.

(iv) **Predictive-based application services** have the highest technical complexity and the greatest business value. For example, personal retail spending records reflect lifestyle habits and can predict some potential health risks, especially diet-related illnesses such as obesity and hypertension. In addition, according to individual physical characteristics, personality characteristics and other factors can be used to predict individual preferences and develop medical plans to meet various needs. For example, Yin Zhang et al. [70] developed iDoctor, a new healthcare referral system based on a hybrid matrix factorization approach. iDoctor predicts users' sentiments and preferences by mining user reviews and evaluations of physicians, thereby providing users with specialized, personalized doctor referrals. iDoctor improves the accuracy of healthcare advice significantly by providing a higher forecast rating.

7. Applications

7.1. Medical Recommendations. In healthcare, the development of the vulnerability to a disease is often considered permanent, but some patients may change from week to week and even become robust again. However, once established, vulnerability is almost impossible to reverse, and less than 1% of hospitalized patients were found to have returned during a five-year follow-up period. Hospital readmission, medical costs, institutionalization, and mortality rates will be greatly enhanced. Therefore, we need to promote positive and healthy aging and incitement measures to prevent vulnerability. From a practical point of view, targeting fragility represents a reasonable method. In particular, the use of multivariate interventions to screen, monitor, and manage prefragility-associated precursors, such as subjective or mild cognitive impairment, can be effectively achieved by mobile medical treatments. Corresponding mobile medical equipment [71], considering the timely diagnosis of the patient's physical condition and reflections, can effectively avoid numerous accidents, reducing the patient's number of adverse health outcomes.

Today, data-driven thinking and methods play a key role in the emergence of personalized medicine. Many diseases have preventable risk factors or at least are dangerous. Clarifying these disease characteristics may help to not only improve personalized healthcare but also reduce the burden of disease. However, the combination of possible risk factors is so complex that it is impossible for an individual physician to analyze it completely (in real time) during patient interactions. Currently, a provider will carefully examine the patient's medical history and perform physical examinations and selective laboratory tests to determine the patient's health condition and future disease risk. These diseases are usually confined to a few diseases, as well as the skills and knowledge of individual providers and the priorities defining individual visits. Thus, taking the next step in personalized healthcare requires the calculation and analysis of big data aggregation and integration frameworks, discovering deep insights into patient similarities and connections, and providing personalized disease risk profiles for each patient's health in a summarized manner[72].

With the increasing use of contemporary mobile messaging data, the mobility of communications technologies and information is also being greatly enhanced. Mobile healthcare, called mobile health or mHealth, has drawn the attention of many practitioners, researchers, and policymakers. Mobile health has the potential to revolutionize healthcare, especially in countries with inadequate medical infrastructure and services in low-income and middle-income low-resource settings [73].

In [70], mixing matrix factorization-based medical recommendations are proposed; advice based on mining the user's evaluations and judgment on the doctor's emotion and preferences developed to provide users with professional, personalized medical treatment is recommended. Specifically, the proposed scheme makes the following contributions to intelligent healthcare services:

(i) A sentiment analysis module, which can calculate a user's emotional state.

(ii) A topic modeling module, which is used to extract the distribution of user preferences and doctor features.

(iii) A hybrid matrix factorization module, which is integrated with two feature distributions extracted by LDA for rating prediction.

7.2. Disease Detection Assisted by Data Analytics. The amount of health-related data has risen sharply in recent years. It is also worth mentioning that the medical insurance reimbursement model is changing, and in today's healthcare environment, meaningful use and performance pay are becoming important new factors. Although profits are not and should not be the main incentive factors, for medical institutions, it is essential to obtain available tools, infrastructure, and technology that can effectively use big data; otherwise, revenues and profits may decrease substantially. Big data includes a variety of features, such as diversity and speed, and it has specific requirements for healthcare. Existing analytical techniques can be applied to a large number of currently unanalyzed health and health data related to patients to better understand the results and then apply them in nursing. Ideally, personal data will inform every doctor and their

patient in the decision-making process and help determine the most appropriate treatment plan [31].

In recent years, the healthcare system in the United States has been rapidly adopting electronic health records, which will greatly increase the amount of clinical data available electronically. Simultaneously, rapid progress has also been made in clinical analysis techniques to analyze large amounts of data and to gather new insights from the analyses. As a result, we have unprecedented opportunities to reduce the cost of healthcare by using mass data. Here are six key use cases that can reduce costs using big data: high-cost patients, readmissions, shunts, decompensation (when the patient's condition is deteriorating), adverse events, and treatment to optimize conditions affecting multiple organ systems. We discuss the type of opinions that may be obtained from the clinical analysis of the types of data that are required to obtain these insights as well as the infrastructure analysis, algorithms, registries, assessment scores, monitoring equipment, etc. Organizations will need to undertake the necessary analysis and implement improvement measures to improve care and reduce costs [74].

Numerous studies have shown that big data analysis has great potential for improving patient care. However, the use of big data in healthcare is still in its infancy and evidence to date suggests that big data analysis will improve outcomes of care to a minimal extent. However, if big data analysis shows improved quality of care and patient outcomes and can be successfully implemented in practice, big data will see its full potential as a significant component of learning in a healthcare system [75].

7.3. Wearable Healthcare Systems. The 21st century is the information age. In a few short years, with the development of 3G and 4G, the progress of information technology has had a tremendous impact on all walks of life. The use of smartphones and tablets has changed the fields of communications, commerce, and entertainment. The technology is changing the way healthcare is delivered, including the quality of patient experience and the cost of healthcare. Mobile technology is improving chronic disease management, empowering the elderly and expectant mothers, reminding people to provide services to underserved areas, and improving the health and efficiency of the healthcare system when it is most needed [76].

To better integrate mobile healthcare into our daily life, we adopt wearable devices, wearable sensor systems are likely to produce more than we can currently easily organize, and our ability to explain these dataset is insufficient. To successfully use wearable sensor data to estimate health status and realize improved health management, we must set standards and ontologies between study groups and business systems to share data and promote the integration of these data and health information systems. However, policies and regulations will need to ensure that the details of wearable sensor data are not abused to violate individual privacy or discriminate [77].

To make mobile healthcare easier, more functional, and more comfortable, we consider flexible and scalable sensors that have recently become an active area of research in wearable, implantable, and resorbable systems for mobile health (mHealth) for achieving extensive and unobtrusive health monitoring. Despite this, there is a lack of systematic research comparing the performance of these new sensors to that of conventional sensors. A novel technique that guides the future design of sensors by printing serpentine, flexible, and retractable electrodes that are attached to the skin during the transfer process and using area density (AD) as a key parameter has been developed. These sensors are used to capture an electrophysiological signal, the electrocardiogram (ECG), and are different from when the ECG is obtained with conventional gold and stainless steel metal clips. As a result of this study, flexible sensors designed for larger capture areas yield higher signal-to-noise ratios (SNRs). In particular, ECGs comparable to conventional metal clips (SNR of 25 dB) can be achieved with this new flexible sensor design, while the new flexible sensor has a design value of 40%. Thus, this new wearable and flexible electrode design can be used not only for human sensing but also for internal measurements of the gastrointestinal tract [78].

In general, today's wearable health monitoring systems may include various types of microsensors, wearable devices, and even implants. These biosensors are capable of measuring vital physiological parameters such as heart rate, blood pressure, body and skin temperature, oxygen saturation, respiratory rate, and electrocardiogram. Measurements are made by wireless or wired connections to a central node such as a personal digital assistant (PDA) or board; then, they are displayed on a user interface or disseminated to the medical center as aggregated vital signs. In our example, we illustrate the fact that wearable medical systems may include a wide variety of components: sensors, wearable materials, smart textiles, actuators, power supplies, wireless communication modules and links, control and processing units, user interfaces, software, and advanced algorithms for data extraction and decision making [79].

Finally, after a research study, we found Wearable 2.0 to be a better choice for today's mobile healthcare. The Wearable 2.0 front-end system includes a wide range of sensors and serves as a data source for the long-term collection of data that are important to health. Moreover, its front-end system can also function as a user interface. In addition, to achieve a good user experience, medical robots can be presented in the implementation of the front-end system. In particular, mobile robots and human enclosures can provide friendlier and more personalized medical services. For example, when an individual suffers a heart attack and the user loses verbal ability, a medical robot can send a video recording and a picture to a telemedicine center or immediate family member. In addition, walking-capable humanoid games play an important role in emotional interaction when emotion sensing services are required. In this way, the integration of smart garments and humanoid robots helps to improve the interoperability of the system in a variety of complex situations. With the support of mobile cloud systems, big data analysis of healthcare big data during long-term storage and in the cloud can greatly enhance the intelligence and awareness of humanoid robots. Therefore, real-time emotional

human-computer interaction can be used with human-robot-based support to provide a certain understanding of the user's intentions by the user. Moreover, a robot can also collect environmental data as a mobile receiver. In short, smart garments support high mobility, while robots provide efficient data sensing and health monitoring. Thus, from a number of perspectives, Wearable 2.0 is our best choice for researching mobile healthcare.

8. Conclusion

With the development of data analytics and mobile computing, healthcare systems are able to provide more intelligent and convenient applications and services. Moreover, assisted by machine learning, data mining, artificial intelligence, and other advanced techniques, healthcare systems could also play an important role as a guide of healthy lifestyles, as a tool to support decision making, and as a source of innovation in the evolving healthcare ecosystem. This paper presents the intelligent healthcare systems assisted by data analytics and mobile computing, therein consisting of the data collection layer, the data management layer, and the service layer. This paper also introduces some representative applications based on the proposed scheme, which have been proved or demonstrated to be able to provide more intelligent, professional, and personalized healthcare services.

Although this paper presents a comprehensive system design for intelligent healthcare systems assisted by data analytics and mobile computing, more advanced techniques should be included in our future work, such as cognitive computing, deep learning, and affective computing, to further improve the quality of service and user experience.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the China National Natural Science Foundation under Grant 61702553 and the Project of Humanities and Social Sciences (17YJCZH252) funded by the China Ministry of Education (MOE).

References

[1] M. Chen, J. Yang, Y. Hao, S. Mao, and K. Hwang, "A 5G Cognitive System for Healthcare," *Big Data and Cognitive Computing*, vol. 1, no. 1, p. 2, 2017.

[2] M. M. Hassan, K. Lin, X. Yue, and J. Wan, "A multimedia healthcare data sharing approach through cloud-based body area network," *Future Generation Computer Systems*, vol. 66, pp. 48–58, 2017.

[3] A. M. Rahmani, T. N. Gia, B. Negash et al., "Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach," *Future Generation Computer Systems*, vol. 78, pp. 641–658, 2018.

[4] M. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C.-H. Youn, "Wearable 2.0: Enabling Human-Cloud Integration in Next Generation Healthcare Systems," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 54–61, 2017.

[5] P.-L. Wu, W. Kang, A. Al-Nayeem, L. Sha, R. B. Berlin Jr., and J. M. Goldman, "A low complexity coordination architecture for networked supervisory medical systems," in *Proceedings of the 4th ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS 2013*, pp. 89–98, usa, April 2013.

[6] G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," *Communications of the ACM*, vol. 57, no. 3, pp. 78–85, 2014.

[7] K. Lee, T. T. H. Wan, and H. Kwon, "The relationship between healthcare information system and cost in hospital," *Personal and Ubiquitous Computing*, vol. 17, no. 7, pp. 1395–1400, 2013.

[8] M. Chen, "NDNC-BAN: supporting rich media healthcare services via named data networking in cloud-assisted wireless body area networks," *Information Sciences*, vol. 284, no. 10, pp. 142–156, 2014.

[9] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[10] H. Zhang, S. Mehotra, D. Liebovitz, C. A. Gunter, and B. Malin, "Mining deviations from patient care pathways via electronic medical record system audits," *ACM Transactions on Management Information Systems (TMIS)*, vol. 4, no. 4, 2013.

[11] R. Mehmood, M. A. Faisal, and S. Altowaijri, "Future networked healthcare systems: A review and case study," *Handbook of Research on Redesigning the Future of Internet Architectures*, pp. 531-55, 2015.

[12] M. I. Pramanik, R. Y. K. Lau, H. Demirkan, and M. A. K. Azad, "Smart health: Big data enabled health paradigm within smart cities," *Expert Systems with Applications*, vol. 87, pp. 370–383, 2017.

[13] M. M. Rathore, A. Paul, A. Ahmad, M. Anisetti, and G. Jeon, "Hadoop-Based Intelligent Care System (HICS)," *ACM Transactions on Internet Technology (TOIT)*, vol. 18, no. 1, pp. 1–24, 2017.

[14] S. V. B. Peddi, P. Kuhad, A. Yassine, P. Pouladzadeh, S. Shirmohammadi, and A. A. N. Shirehjini, "An intelligent cloud-based data processing broker for mobile e-health multimedia applications," *Future Generation Computer Systems*, vol. 66, pp. 71–86, 2017.

[15] R. Nambiar, R. Bhardwaj, A. Sethi, and R. Vargheese, "A look at challenges and opportunities of Big Data analytics in healthcare," in *Proceedings of the 2013 IEEE International Conference on Big Data, Big Data 2013*, pp. 17–22, usa, October 2013.

[16] J. P. McGlothlin and L. Khan, "Managing evolving code sets and inte- gration of multiple data sources in health care analytics," in *Proceedings of the in Proceedings of the 2013 international workshop on Data management analytics for healthcare*, p. 14, 2013.

[17] V. Chandola, S. R. Sukumar, and J. Schryver, "Knowledge discovery from massive healthcare claims data," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2013*, pp. 1312–1320, usa, August 2013.

[18] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Information Fusion*, vol. 35, pp. 1339–1351, 2017.

[19] P.-Y. Wu, C.-W. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, and M. D. Wang, "-Omic and Electronic Health Record Big Data Analytics for Precision Medicine," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, pp. 263–273, 2017.

[20] C.-H. Lin, L.-C. Huang, S.-C. T. Chou, C.-H. Liu, H.-F. Cheng, and I.-J. Chiang, "Temporal event tracing on big healthcare data analytics," in *Proceedings of the 3rd IEEE International Congress on Big Data, BigData Congress 2014*, pp. 281–287, usa, July 2014.

[21] M. G. Rabiul Alam, E. J. Cho, E.-N. Huh, and C. S. Hong, "Cloud based mental state monitoring system for suicide risk reconnaissance using wearable bio-sensors," in *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2014*, khm, January 2014.

[22] C. He, X. Fan, and Y. Li, "Toward ubiquitous healthcare services with a novel efficient cloud platform," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 230–234, 2013.

[23] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: A survey of the state of the art and research challenges," *ACM Computing Surveys*, vol. 47, no. 3, 2015.

[24] J. Gikas and M. M. Grant, "Mobile computing devices in higher education: Student perspectives on learning with cellphones smartphones & social media," *Internet & Higher Education*, vol. 19, pp. 18–26, 2013.

[25] C. L. Ventola, "Mobile devices and apps for health care professionals: Uses and benefits," *Pharmacy Therapeutics*, vol. 39, no. 5, pp. 356-64, 2014.

[26] S. Kumar, W. Nilsen, M. Pavel, and M. Srivastava, "Mobile health: revolutionizing healthcare through transdisciplinary research," *The Computer Journal*, vol. 46, no. 1, Article ID 6357165, pp. 28–35, 2013.

[27] Y.-H. Kao, B. Krishnamachari, M.-R. Ra, and F. Bai, "Hermes: Latency optimal task assignment for resource-constrained mobile computing," in *Proceedings of the 34th IEEE Annual Conference on Computer Communications and Networks, IEEE INFOCOM 2015*, pp. 1894–1902, hkg, May 2015.

[28] S. Altowaijri, R. Mehmood, and J. Williams, "A quantitative model of grid systems performance in healthcare organisations," in *Proceedings of the 2010 International Conference on Intelligent Systems, Modelling and Simulation*, pp. 431–436, January 2010.

[29] M. Viceconti, P. Hunter, and R. Hose, "Big Data, Big Knowledge: Big Data for Personalized Healthcare," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1209–1215, 2015.

[30] Z. Lv, J. Chirivella, and P. Gagliardo, "Bigdata oriented multimedia mobile health applications," *Journal of Medical Systems*, vol. 40, no. 5, pp. 1–10, 2016.

[31] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science & Systems*, vol. 2, no. 1, p. 3, 2014.

[32] A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," *BioMed Research International*, vol. 2015, Article ID 370194, 2015.

[33] C.-J. Su and C.-Y. Chiang, "Iaserv: An intelligent home care web services platform in a cloud for aging-in-place," *International Journal of Environmental Research and Public Health*, vol. 10, no. 11, pp. 6106–6130, 2013.

[34] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: review and open research issues," *Information Systems*, vol. 47, pp. 98–115, 2015.

[35] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2017.

[36] L. A. Tawalbeh, R. Mehmood, E. Benkhlifa, and H. Song, "Mobile Cloud Computing Model and Big Data Analysis for Healthcare Applications," *IEEE Access*, vol. 4, pp. 6171–6180, 2016.

[37] J. Wan, C. Zou, S. Ullah, C.-F. Lai, M. Zhou, and X. Wang, "Cloud-Enabled wireless body area networks for pervasive healthcare," *IEEE Network*, vol. 27, no. 5, pp. 56–61, 2013.

[38] N. Sultan, "Making use of cloud computing for healthcare provision: opportunities and challenges," *International Journal of Information Management*, vol. 34, no. 2, pp. 177–184, 2014.

[39] A. Grunerbl, A. Muaremi, V. Osmani et al., "Smart-phone based recognition of states and state changes in bipolar disorder patients," *Biomedical & Health Informatics IEEE Journal*, vol. 19, no. 1, pp. 140–148, 2015.

[40] M. M. Rodgers, V. M. Pai, and R. S. Conroy, "Recent advances in wearable sensors for health monitoring," *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3119–3126, 2015.

[41] M. Chen, Y. Zhang, Y. Li, M. M. Hassan, and A. Alamri, "AIWAC: affective interaction through wearable computing and cloud technology," *IEEE Wireless Communications Magazine*, vol. 22, no. 1, pp. 20–27, 2015.

[42] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 522–531, 2015.

[43] A. Pyattaev, K. Johnsson, S. Andreev, and Y. Koucheryavy, "Communication challenges in high-density deployments of wearable wireless devices," *IEEE Wireless Communications Magazine*, vol. 22, no. 1, pp. 12–18, 2015.

[44] H. Ghasemzadeh, P. Panuccio, S. Trovato, G. Fortino, and R. Jafari, "Power-aware activity monitoring using distributed wearable sensors," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 537–544, 2014.

[45] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco, "RFID technology for IoT-based personal healthcare in smart spaces," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 144–152, 2014.

[46] L. Catarinucci, D. de Donno, L. Mainetti et al., "An IoT-aware architecture for smart healthcare systems," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 515–526, 2015.

[47] B. Y. Xu, L. D. Xu, H. M. Cai, C. Xie, J. Y. Hu, and F. Bu, "Ubiquitous data accessing method in iot-based information system for emergency medical services," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1578–1586, 2014.

[48] G. Yang, L. Xie, M. Mäntysalo et al., "A health-IoT platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2180–2191, 2014.

[49] S. A. Haque, S. M. Aziz, and M. Rahman, "Review of cyber-physical system in healthcare," *International Journal of Distributed Sensor Networks*, vol. 2014, Article ID 217415, 2014.

[50] X. Hu, T. H. S. Chu, H. C. B. Chan, and V. C. M. Leung, "Vita: a crowdsensing-oriented mobile cyber-physical system," *IEEE Transactions on Emerging Topics in Computing*, vol. 1, no. 1, pp. 148–165, 2013.

[51] A. Costanzo, A. Faro, D. Giordano, and C. Pino, "Mobile cyber physical systems for health care: Functions, ambient ontology and e-diagnostics," in *Proceedings of the 13th IEEE Annual Consumer Communications and Networking Conference, CCNC 2016*, pp. 972–975, usa, January 2016.

[52] J. Schobel, R. Pryss, M. Schickler, and M. Reichert, "Towards flexible mobile data collection in healthcare," in *Proceedings of the 29th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2016*, pp. 181-182, irl, June 2016.

[53] P. D. Kaur and I. Chana, "Cloud based intelligent system for delivering health care as a service," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 346–359, 2014.

[54] A. Gaggioli, G. Pioggia, G. Tartarisco et al., "A mobile data collection platform for mental health research," *Personal and Ubiquitous Computing*, vol. 17, no. 2, pp. 241–251, 2013.

[55] M. Chen, S. Mao, Y. Zhang, and V. C. Leung, "Related Technologies," in *Big Data*, SpringerBriefs in Computer Science, pp. 11–18, Springer International Publishing, Cham, 2014.

[56] G. DeCandia, D. Hastorun, M. Jampani et al., "Dynamo: amazon's highly available key-value store," in *Proceedings of the 21st ACM Symposium on Operating Systems Principles (SOSP '07)*, pp. 205–220, ACM, October 2007.

[57] J. Han, E. Haihong, G. Le, and J. Du, "Survey on NoSQL database," in *Proceedings of the 6th International Conference on Pervasive Computing and Applications (ICPCA '11)*, pp. 363–366, Port Elizabeth, South Africa, October 2011.

[58] L. George, *HBase - The Definitive Guide: Random Access to Your Planet-Size Data*, DBLP, 2011.

[59] K. Chodorow and M. Dirolf, *MongoDB: The Definitive Guide Powerful and Scalable Data Storage*, DBLP, 2010.

[60] D. Borthakur, *Hdfs Architecture Guide*, 2008.

[61] J. C. Corbett, J. Dean, M. Epstein et al., "Spanner: Googles globally-distributed database," in *Proceedings of the Usenix Conference on Op- erating Systems Design and Implementation*, pp. 251–264, 2012.

[62] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop distributed file system," in *Proceedings of the IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST '10)*, 10, 1 pages, Piscataway, NJ, USA, May 2010.

[63] J. G. Shanahan and L. Dai, "Large scale distributed data science using apache spark," in *Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2015*, pp. 2323-2324, aus, August 2015.

[64] D. P. Anderson, "BOINC: a system for public-resource computing and storage," in *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*, pp. 4–10, IEEE, November 2004.

[65] A. G. Shoro and T. R. Soomro, "Big data analysis: Apache spark perspective," *Big data analysis: Apache spark perspective*, vol. 15, 2015.

[66] M. C. De, E. Kiciman, M. Dredze et al., "Discovering shifts to suicidal ideation from mental health content," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 2098–2110, 2016.

[67] G. Fortino, D. Parisi, V. Pirrone, and G. Di Fatta, "BodyCloud: a SaaS approach for community body sensor networks," *Future Generation Computer Systems*, vol. 35, pp. 62–79, 2014.

[68] À. Bravo, J. Piñero, N. Queralt-Rosinach, M. Rautschka, and L. I. Furlong, "Extraction of relations between genes and diseases from text and large-scale data analysis: Implications for translational research," *BMC Bioinformatics*, vol. 16, no. 1, article no. 55, 2015.

[69] H. Lu, B. Li, J. Zhu et al., "Wound intensity correction and segmentation with convolutional neural networks," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 6, p. e3927, 2017.

[70] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "IDoctor: personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generation Computer Systems*, vol. 66, pp. 30–35, 2017.

[71] R. O'Caoimh, D. W. Molloy, C. Fitzgerald et al., "Healthcare recommendations from the personalised ict supported service for independent living and active ageing (PERSSILAA) study," in *Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health, ICT4AWE 2017*, pp. 91–103, prt, April 2017.

[72] F. F. Costa, "Big data in biomedicine," *Drug Discovery Therapy*, vol. 19, no. 4, pp. 433–440, 2014.

[73] A. Chib, M. H. Van Velthoven, and J. Car, "MHealth adoption in low-resource environments: A review of the use of mobile healthcare in developing countries," *Journal of Health Communication*, vol. 20, no. 1, pp. 4–34, 2015.

[74] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.

[75] J. S. Rumsfeld, K. E. Joynt, and T. M. Maddox, "Big data analytics to improve cardiovascular care: Promise and challenges," *Nature Reviews Cardiology*, vol. 13, no. 6, pp. 350–359, 2016.

[76] D. West, *How Mobile Devices Are Transforming Healthcare How Mobile Devices Are Transforming Healthcare*, Washington, Wash, USA, 2012.

[77] S. J. Redmond, N. H. Lovell, G. Z. Yang et al., "What Does Big Data Mean for Wearable Sensor Systems? Contribution of the IMIA Wearable Sensors in Healthcare WG," *Yearbook of Medical Informatics*, vol. 9, pp. 135–142, 2014.

[78] N. Luo, J. Ding, N. Zhao, B. H. K. Leung, and C. C. Y. Poon, "Mobile health: Design of flexible and stretchable electrophysiological sensors for wearable healthcare systems," in *Proceedings of the 11th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2014*, pp. 87–91, che, June 2014.

[79] A. Pantelopoulos and N. G. Bourbakis, "A survey on wearable sensor- based systems for health onitoring and prognosis," *IEEE Transactions on Systems Man Cybernetics Part C*, vol. 40, p. 12, 2009.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5412925, 10 pages
https://doi.org/10.1155/2018/5412925

Research Article

A Mobile Computing Method Using CNN and SR for Signature Authentication with Contour Damage and Light Distortion

Mei Wang[1], **Ke Zhai**[1], **Chi Harold Liu**[2,3] **and Yujie Li**[4]

[1]*School of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an, China*
[2]*School of Software, Beijing Institute of Technology, China*
[3]*Department of Computer Information and Security, Sejong University, Republic of Korea*
[4]*School of Information Engineering, Yangzhou University, Yangzhou 225127, China*

Correspondence should be addressed to Mei Wang; wangm@xust.edu.cn and Yujie Li; yzyjli@gmail.com

Received 21 November 2017; Accepted 21 March 2018; Published 25 June 2018

Academic Editor: Haider Abbas

A signature is a useful human feature in our society, and determining the genuineness of a signature is very important. A signature image is typically analyzed for its genuineness classification; however, increasing classification accuracy while decreasing computation time is difficult. Many factors affect image quality and the genuineness classification, such as contour damage and light distortion or the classification algorithm. To this end, we propose a mobile computing method of signature image authentication (SIA) with improved recognition accuracy and reduced computation time. We demonstrate theoretically and experimentally that the proposed golden global-local (G-L) algorithm has the best filtering result compared with the methods of mean filtering, medium filtering, and Gaussian filtering. The developed minimum probability threshold (MPT) algorithm produces the best segmentation result with minimum error compared with methods of maximum entropy and iterative segmentation. In addition, the designed convolutional neural network (CNN) solves the light distortion problem for detailed frame feature extraction of a signature image. Finally, the proposed SIA algorithm achieves the best signature authentication accuracy compared with CNN and sparse representation, and computation times are competitive. Thus, the proposed SIA algorithm can be easily implemented in a mobile phone.

1. Introduction

Artificial intelligence influences the information technology being developed in today's world. People use artificial intelligence digital information technology almost anywhere and at any time. This supports daily social life and economic activities and contributes greatly to the sustainable growth of the economy and solves various social problems [1]. A signature is a commonly used human feature for identity authentication [2–4]. Artificial intelligence approaches to signature authentication have been evolving from offline methods to online methods to meet modern demands.

Regarding offline methods, researchers have developed signature recognition methods using a fusion algorithm involving distance and centroid orientation [5]. Additionally, the single optimized dehazing method has been proposed to estimate atmospheric light and remove the haze from an image [6], and methods to ensure consistency of signature verification have been proposed [7–10]. Dissimilarity normalization, shape features, and complex network spectrums have also been developed to assist in signature verification [11–14]. In addition, scientists presented a multitask metric learning recognition method that uses true and fake samples to calculate the similarity and dissimilarity of a signature [15]. Fine geometric structures were encoded by using a mesh template and splitting the area of the subset of features to analyze and verify a signature [16, 17].

Regarding online methods, scientists have utilized fast Fourier transform (FFT) [18] and dynamic time warping (DTW) [19] methods for on-air signature verification based on video, leading to the discrete cosine transform developed for online signature verification [20]. Researchers have also developed a signature alignment verification method to obtain a best match effect on the basis of a Gaussian mixture

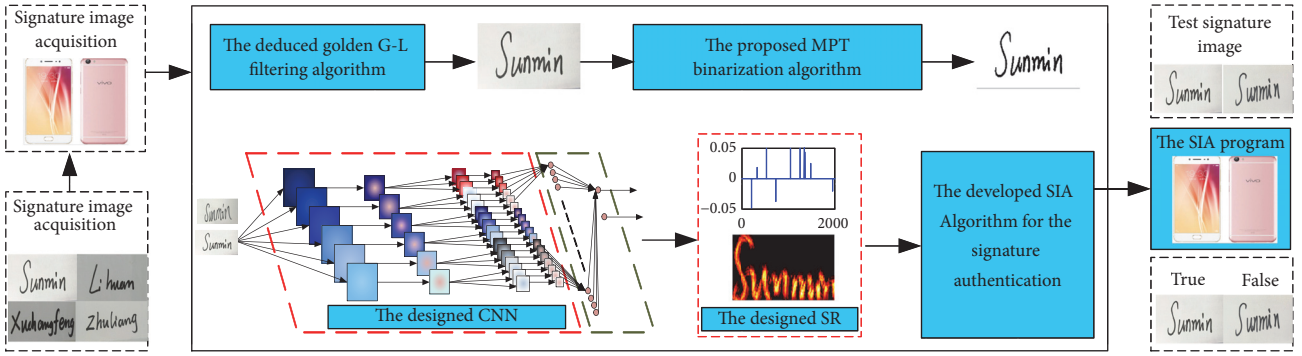


Figure 1: Signature authentication scheme.

algorithm [21]. Scientists have further developed various image-processing approaches, such as image descattering, color restoration, and image quality assessments that support signature verification activities.

Current state-of-the-art methods include deep learning and artificial intelligence. Recently, the convolutional neural network (CNN) method has become a popular research topic in many applications. Additionally, researchers have proposed deep probabilistic neural networks and optimal parameters determined by particle swarm optimization (PSO) as a signature method design. In more modern applications, the traditional template signature has been replaced by the hidden signature to minimize the mean misalignment. Scientists have applied CNN to establish a fast level set algorithm to solve the color distortion problem for wound image intensity correction and segmentation processing, which could have applications in signature verification [22].

Deep CNN was originally developed for object classification, and reinforcement learning was developed to detect abnormal information [23]. For signature image authentication (SIA), scientists have proposed the improved CNN method to process distorted samples and decrease the distortion [24, 25]. Sparse representation (SR) has been proposed for a separate feature-level fusion to integrate multiple feature representation [26]. Also, the local SR was proposed to improve robustness in cases of partial occlusion, deformation, and rotation in visual tracking [27]. The hierarchical SR was developed for synthetic aperture radar image classification [28]. In addition, researchers have developed the probabilistic class structure [29] and the sparse exponent batch processing method [30] combined with artificial intelligence methods for signature verification.

Meanwhile, mobile systems have been developing rapidly, and the above methods have not considered signature authentication applications for mobile systems, especially mobile phones. Based on the lack of such research, this paper addresses the problems of contour damage and light distortion as well as the classification accuracy of signature images. A combination of CNN and SR is proposed as a potential signature authentication method for mobile phones.

The organization of this paper is as follows: Section 2 presents the basic design for the signature authentication system, the golden global-local (G-L) filtering algorithm to solve the contour damage problem, the minimum probability threshold (MPT) segmentation algorithm to obtain the minimum error result, the CNN design for decreasing the light distortion, and the SIA algorithm to increase the recognition accuracy and obtain better speed performance. Section 3 discusses the experiments and comparisons. Finally, Section 4 concludes the paper.

2. Materials And Methods

The proposed system collects signature image samples using a mobile phone with the CamScanner app. Then ACDSee software is used for image cutting creating an image size of 64×128 pixels. Furthermore, MATLAB is used to conduct the CNN training and the SR method design. Finally, the SIA result is obtained and presented as output.

We design the signature authentication system scheme as shown in Figure 1. First, true signature images are collected by the mobile phone. Then, the signature verification system is applied on the basis of the deduced golden G-L filtering algorithm, the proposed MPT algorithm, and the proposed SIA algorithm. Finally, the application program of the SIA is installed in a mobile phone.

2.1. Derivation Of Golden G-L Filtering Algorithm For Contour Damage Mending. In order to mend the contour damage, remove noise, and smooth the signature image, a filtering process is needed first. Filtering is a convolution process of the input signature image with a core. Commonly used filtering methods are Gaussian filtering and mean filtering.

However, the Gaussian filtering effect should be improved, while mean filtering lacks scaling properties in variance and rotation symmetry. Therefore, we developed the golden G-L filtering algorithm below.

An original signature image, $I_0 = f(m \times n)$, occupies the total area $S_0 = (m \times n)$. The global mean gray value M_0 of image I_0 is

$$M_0 = \frac{1}{m \times n} \sum_{(i,j) \in S_0} f(i,j) \tag{1}$$

where f is the gray value of a pixel and (i, j) are the coordinates of a pixel, and m and n are the row number

Table 1: Deduced golden G-L filtering algorithm for mending contour damage of signature image.

Input:	The original signature image $I_0 = f(x, y)$.
Output:	The filtered image $I_1(x, y)$.
Step 1:	For each pixel of I_0, calculate the global mean gray value M_0 and the variance δ^2 according to (1) and (2), respectively.
Step 2:	Calculate the local mean gray value M_1 of the pixel according to (3).
Step 3:	Calculate the G-L mean gray value M_g of the pixel according to (4).
Step 4:	Calculate the improved Gaussian template according to (5) and (6).
Step 5:	Add salt and pepper noise signature image I_0 and obtain the filtered image I_1.
Step 6:	For each pixel, filter the noise signature image I_1 by the convolution operation using the improved Gaussian template $G(x, y)$, and obtain the filtered image I_2.

and the column number of the original signature image I_0, respectively.

The gray value variance δ^2 of the original signature image $I_0 = f(m \times n)$ is

$$\delta^2 = \frac{1}{m \times n - 1} \sum_{k=1}^{m \times n} (M - M_0)^2 \tag{2}$$

where M_0 is the global mean gray value of image I_0, M_1 is the local mean gray value of a pixel, and m and n were defined previously.

In order to obtain a better filtering effect, we define the G-L mean parameter M_g by a combination of the mean filtering method with a golden section number $\sigma = 0.618$.

The local neighbor area S_1 of a pixel is selected to be five rows and five columns, so it occupies the local area $S_0 = (5 \times 5)$. The local mean gray value M_1 of image I_0 is then

$$M_1 = \frac{1}{5 \times 5} \sum_{(i,j) \in S_1} f(p, q) \tag{3}$$

where the pixel coordinates are (p, q) and f is the gray value of a pixel.

According to experiments we conducted, a new parameter M_g of the G-L mean gray value is defined below.

$$M_g(x, y) = (1 - \sigma) M_0 + \sigma M_1 \tag{4}$$

where $x = 1, 2, \ldots, m$, $y = 1, 2, \ldots, n$, σ is the golden section number ($\sigma = 0.618$), S_0 and S_1 are the global area and the local area of a pixel, respectively, and (x, y), (i, j), and (p, q) represent pixel coordinates.

The G-L mean M_g is defined as the weighted sum of the global mean gray value M_0 and the local mean gray value M_1 with the golden section σ, and the local mean gray value M_1 is the main component of the G-L mean M_g.

Thus, we determine the pixel positions $\{(i_0, j_0), \ldots, (i_k, j_k), \ldots\}$ where the gray values are equal to M_g.

$$\begin{aligned} &\underset{(i_k, j_k)}{\operatorname{argmin}} \operatorname{distance}\left[(x, y), (i_k, j_k)\right] \\ &\quad := \left\{(i_k, j_k) \mid \forall (i_k, j_k) : M(i_k, j_k) = M_g\right\} \end{aligned} \tag{5}$$

where (x, y) are the coordinates of a pixel of the original signature image I_0, M is the gray value of a pixel, M_0 is the global mean gray value of image I_0, and (i_k, j_k) are the pixel coordinates where the gray value is equal to M_g.

Then, we design the improved Gaussian template $G(x, y)$ as below.

$$G(x, y) = \frac{1}{4\pi\delta^2} \exp\left(-\frac{(x - i_k)^2 + (y - j_k)^2}{4\delta^2}\right), \quad x = 1, \ldots, m, \; y = 1, \ldots, n \tag{6}$$

Finally, the G-L filtering algorithm can be described on the basis of the defined parameter M_g and the improved Gaussian template $G(x, y)$ as outlined in Table 1.

Note that salt and pepper noise is added to the original signature image I_0 in Step 5. This operation is used for mending any contour damage in the signature.

The deduced golden G-L filtering algorithm has scaling advantages in variance and rotation symmetry as well as the contour damage mending effect. This is because of the combination of the mean filtering, the Gaussian filtering and golden section of the global and local information, and the mending operation using salt and pepper noise, respectively.

2.2. Proposed MPT Algorithm For Minimum Error Segmentation. To achieve the minimum error binary segmentation of a signature image, we propose the MPT algorithm on the basis of signature image analysis and the developed optimal threshold for binary segmentation.

Image segmentation sets all pixel values to 0 or 1 while the pixel positions remain invariant. This operation simplifies postprocessing. Using this technique, the binary threshold influences the result dramatically. For example, Figure 2 shows the gray signature image and the histogram as well as the influence of a threshold on the segmentation results.

For a signature image, the gray degree histogram has two peaks. One is for the background of the signature, and the other is for the signature itself. Between these two peaks, there must be a minimum point where the gray degree value $M_{\min}$ corresponds to the minimum histogram value. We select this minimum gray degree value $M_{\min}$ to be the threshold for binary segmentation of the filtered image $I_2(x, y)$. To ensure the minimum error segmentation of a signature image, the proposed MPT algorithm is described in Table 2.

Table 2: Proposed MPT segmentation algorithm with minimum error.

Input:	The filtered image $I_2(x, y)$.
Output:	The binarized image $I_3(x, y)$.
Step 1:	Calculate the gray degree distribution of the filtered signature image $I_2(x, y)$, and draw the histogram.
Step 2:	Calculate the two gray values G_{max1} and G_{max2}, which correspond to the two maximum probabilities in the histogram.
Step 3:	Seek the gray value $G_T \subset [G_{max1}, G_{max2}]$. G_T corresponds to the minimum probability in the range $[G_{max1}, G_{max2}]$, $G_{max1} < G_T < G_{max2}$.
Step 4:	Segment the filtered image $I_2(x, y)$ based on the threshold G_T and obtain the binary image $I_3(x, y)$.

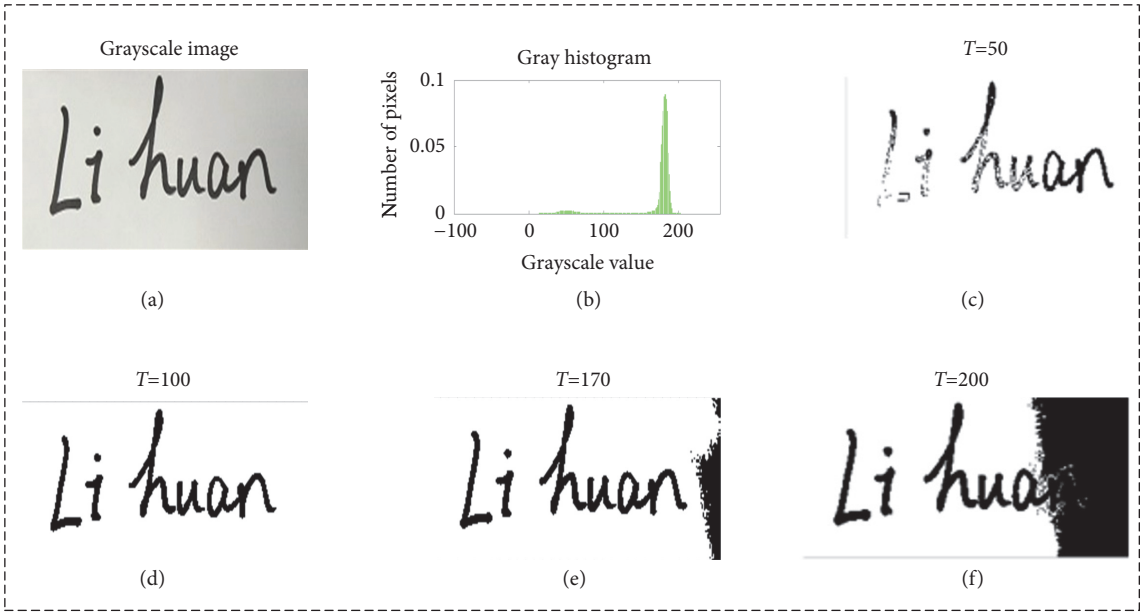


Figure 2: Gray signature image and the histogram program as well as the influence of a threshold on the segmentation result.

2.3. CNN Design For Decreasing Light Distortion. To ignore the light distortion and extract the frame construction and the special details of a signature, a CNN structure is designed, as shown in Figure 3. The CNN has the obvious advantages of local sensing, a hierarchical structure, and integration of the feature extraction and the classification. It is mainly used to verify the two-dimensional graph invariance if displacement, zoom, and other forms of distortion occur. This CNN structure is designed specifically for solving any light distortion problems.

First, the true 64×128 signature image is used as the input for CNN training. Second, we design the six convolution kernels (9×9) for the first stage of feature extraction of the true signature. The convolution layer C_1 is composed of six images (56×120). Third, the sampling kernel S_1 is selected with a size of 2×2 to obtain the pooling layer P_1 which serves as the first part of the inputs to the SR classifier. Then, we design three convolution kernels (5×5) for feature extraction by the second stage.

The pooling layer P_1 is mapped to the convolution layer C_2 to extract the second level of features P_2, which serves as the second part of the inputs of the SR. In the pooling layers P_1 and P_2, the different images focus on different types of features. Some focus on the signature frame, some focus on the key points of a signature, and others focus on the changing areas of a signature. The designed CNN extracts relatively complete features from the original signature features. In this signature authentication system, the image features P_1 and P_2 work together and serve as the inputs of the SR. This CNN is applied to extract the frame construction and the special details of a signature, and eliminates any light distortion problems in a signature image.

2.4. Proposed SIA Algorithm To Increase The Recognition Accuracy. In this section, we design the SIA algorithm based on CNN and SR for signature authentication. The SIA algorithm is described in Table 3.

The SR uses the least number of suitable features for reconstruction of the most complete information. Therefore, the speed of the SR classifier is relatively high. The difficulty with the SR method is to determine the solution of the optimal objective function. Thus, we must solve the two

Table 3: SIA algorithm for signature authentication.

Input:	Training samples of the true signature images, testing samples of the signature images, and the given error parameter ε.
Output:	The recognition result for the unknown signature.
Step 1:	Filter each sample by using the proposed golden G-L filtering algorithm and obtain the sample set Q_1.
Step 2:	Segment each sample of Q_1 using the developed MPT segmentation algorithm and obtain the sample set Q_2.
Step 3:	Design the CNN structure parameters as per Figure 3.
Step 4:	Build the CNN using the sample set Q_2 and obtain the true signature feature images of P_1 and P_2 from the designed CNN.
Step 5:	Calculate the over-complete dictionary D and the sparse coefficient nonzero solutions N by using P_1 and P_2.
Step 6:	Reconstruct the true signature template $SR = D * N$.
Step 7:	Reconstruct the unknown signature TSI using the same method as the true signature template SR.
Step 8:	If $\lvert TSI\text{-}SR\rvert \leqq \varepsilon$, then the test signature is classified as true. Otherwise, the test signature is classified as false.

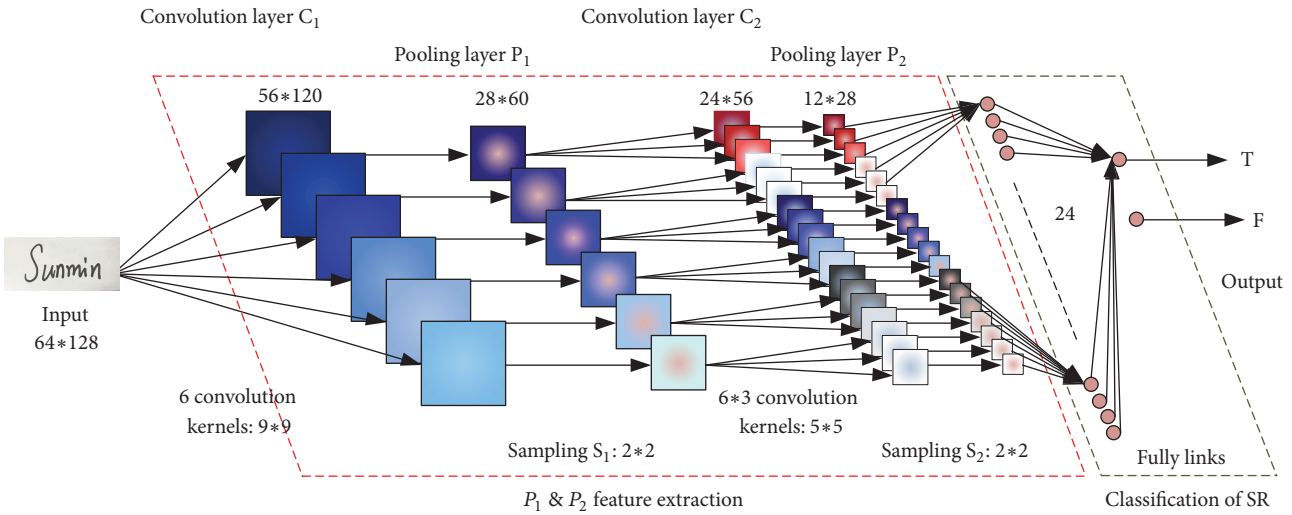


Figure 3: Designed CNN to solve the light distortion problem and extract the detailed frame features of a signature.

problems of obtaining the supercomplete dictionary and the nontrivial solution for sparse coefficients.

After the true signature image template SR is reconstructed, the test signature can be classified as true if the difference between the test signature image *TSI* and the reconstructed true signature image template SR is smaller than the error parameter ε. Otherwise, the test signature is classified as false.

3. Experiments And Analysis

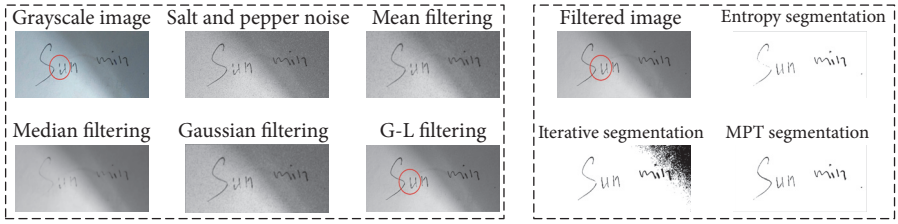
In this section, we demonstrate three experiments of the G-L algorithm, the MPT algorithm, and the SIA algorithm followed by comparisons and discussion. The training and testing datasets of signatures are collected from 300 students. The students include 150 males and 150 females. The true signature number is 300 and the false signature number is also 300.

3.1. G-L Algorithm and MPT Algorithm Experiments. In Section 2.1, a new filtering algorithm, G-L, is developed, and the novel segmentation algorithm MPT is proposed in Section 2.2. Figure 4 shows the original signature image of the true signature and the fake signature. Figure 5 presents comparisons of the G-L filtering algorithm and the MPT segmentation algorithm to traditional methods.

The developed G-L filtering algorithm with added salt and pepper noise has the best filtering effect compared to traditional mean filtering, medium filtering, and Gaussian filtering. The developed G-L filtering algorithm decreases the influence of signature contour damage.

(a) (b)

Figure 4: Original signature images: (a) true and (b) fake.



(a) (b)

Figure 5: Experimental result comparisons of traditional methods and developed methods: (a) the filtering and (b) the segmentation.

The proposed MPT segmentation algorithm produces a better segmentation effect than the traditional methods of maximum entropy and iterative segmentation and has the minimum segmentation error.

3.2. SIA Algorithm Experiment. For the new SIA signature authentication algorithm, 90 true signature images and 72 false signature images for each user are considered. Two-thirds of the samples of true and false images are used for CNN training, and the remaining one-third samples are used for CNN testing. Then, $P1$ and $P2$ obtained from the CNN are selected as the inputs of the SR. Figure 6 shows the signature features P_1 and P_2. P_1, and P_2 retain the frame features and the special detail features of the signature, ignoring any light distortion of the signature image.

For the SR of the SIA, the overcomplete dictionary D and the sparse coefficients N are essential. They are shown in Figures 7 and 8, respectively.

The process and result interface of the signature recognition system is shown in Figure 9. First, the CNN is trained using the training samples and the test samples. Then, the six signature frame features $P1$ and the 18 detailed features $P2$ are obtained, and the total 24 image features (which avoid the light distortion problem) serve as the inputs of the SR.

Finally, the dictionary D and the sparse coefficients N of the true signature are calculated, and the true signature template SR can be reconstructed. In Figure 9, a signature is input to the system to be judged as true or false by the proposed SIA algorithm. The right part shows the sparse coefficients of the 8th chancel of the SR, the central part is the first 36 dictionaries of the SIZE, and the left pan shows the signature authentication result.

3.3. Comparisons and Discussion. Based on the theoretical analysis and the experiments, the comparison of traditional filtering methods and the developed golden G-L filtering algorithm is given in Table 4, the comparison of traditional segmentation methods and the developed MPT segmentation algorithm is given in Table 5, and the performance comparison of the traditional signature authentication methods and the proposed SIA algorithm is given in Table 6. The performance comparisons are shown in Table 7.

From Table 4, the proposed golden G-L filtering algorithm has the best signature contour damage mending and filtering result compared with the traditional methods of mean filtering, medium filtering, and Gaussian filtering. From Table 5, the developed MPT segmentation algorithm has the minimum segmentation error and produces the best signature segmentation compared with the traditional maximum entropy and iterative methods. From Table 6, the performance comparison results indicate that the proposed SIA algorithm has the highest signature authentication accuracy and acceptable time consumption performance compared with the traditional single CNN method and single SR method.

4. Conclusions

It is theoretically and experimentally verified that the proposed golden G-L algorithm has the best filtering result compared with the traditional methods of mean filtering, medium filtering, and Gaussian filtering in the case where the original signature contour is damaged. Meanwhile, the developed MPT algorithm has the best segmentation results with minimum error compared with the maximum entropy and iterative segmentation methods. In addition, the designed CNN can solve the light distortion problem for the feature extraction of the frame features and the detailed features of signature images. Finally, the proposed SIA algorithm achieves the highest average signature authentication accuracy of 97%. In contrast, the average accuracies of the single CNN method and the single SR method are 95% and 94%, respectively. Consumption times are 0.8, 1.0, and 0.7 s, respective to the proposed SIA, CNN, and SR. Future work will focus

Table 4: Signature authentication comparisons.

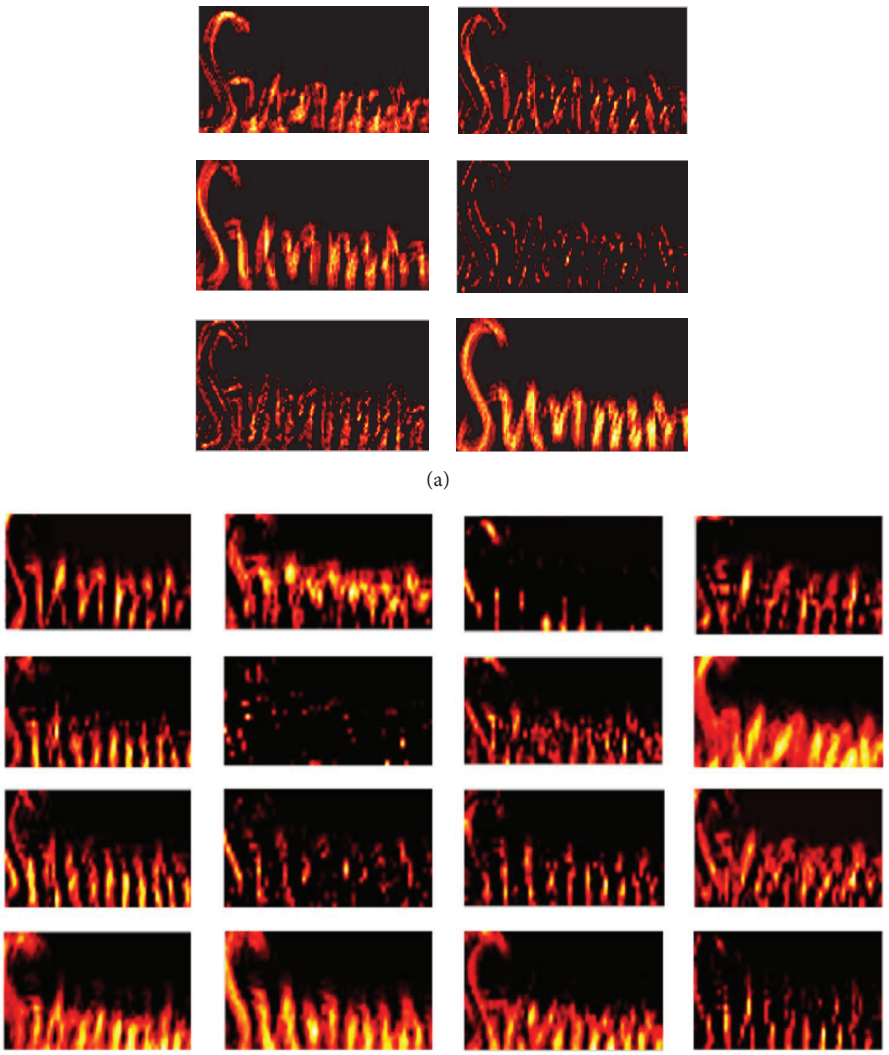
Filtering algorithm	Contour damage mending	Filtering result for signature
Mean filtering	No	Better
Medium filtering	No	Ordinary
Gaussian filtering	No	Good
G-L filtering	Yes	Best

Table 5: Signature authentication comparisons.

Segmentation algorithm	Error rate	Segmentation effect
Maximum entropy	Modest	Better
Iterative method	Maximum	Ordinary
MPT algorithm	Minimum	Best

Table 6: Comparison of signature authentication algorithms.

Authentication algorithm	Accuracy	Time consumption
CNN	Higher	Modest
RS	Modest	Shortest
SIA algorithm	Highest	Short



(a)

(b)

Figure 6: Image features obtained from the designed CNN: (a) P_1 and (b) P_2.

Table 7: Performance comparisons.

Test No.	CNN		RS		SIA	
	A(%)	T(s)	A(%)	T(s)	A(%)	T(s)
1	96	0.8	93	0.6	99	0.6
2	95	0.9	87	0.9	95	1.1
3	94	1.0	95	0.5	99	0.7
4	92	1.2	92	0.7	96	0.9
5	93	1.1	91	0.8	96	0.7
Mean value	95	1.0	94	0.7	97	0.8

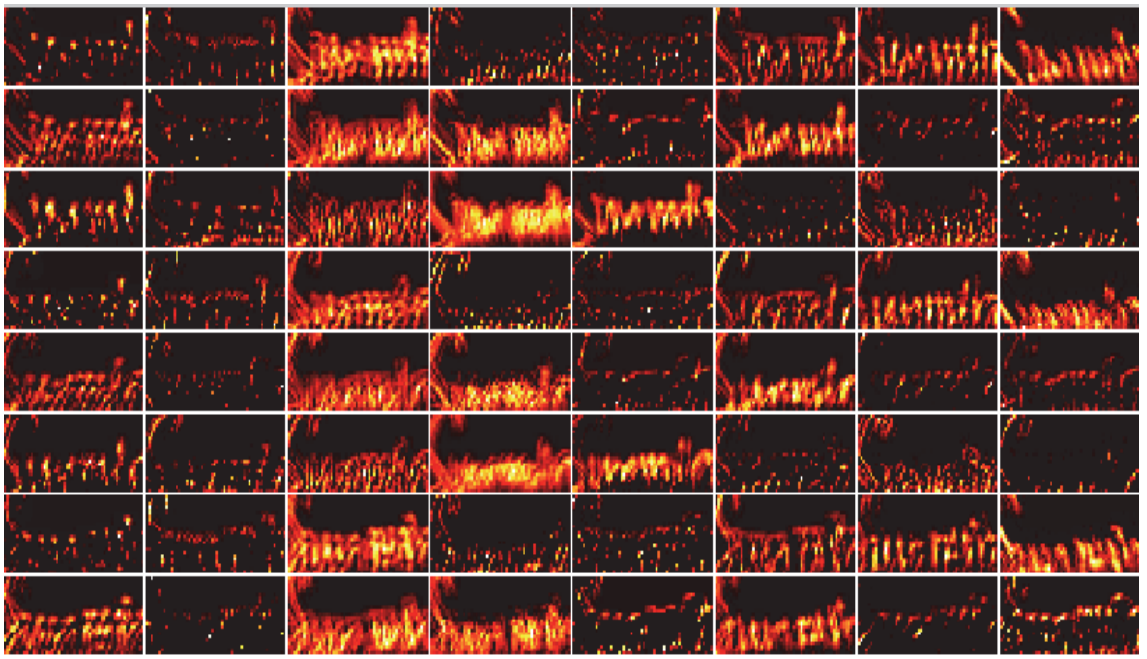


Figure 7: The first 64 overcomplete dictionaries D of the SR.

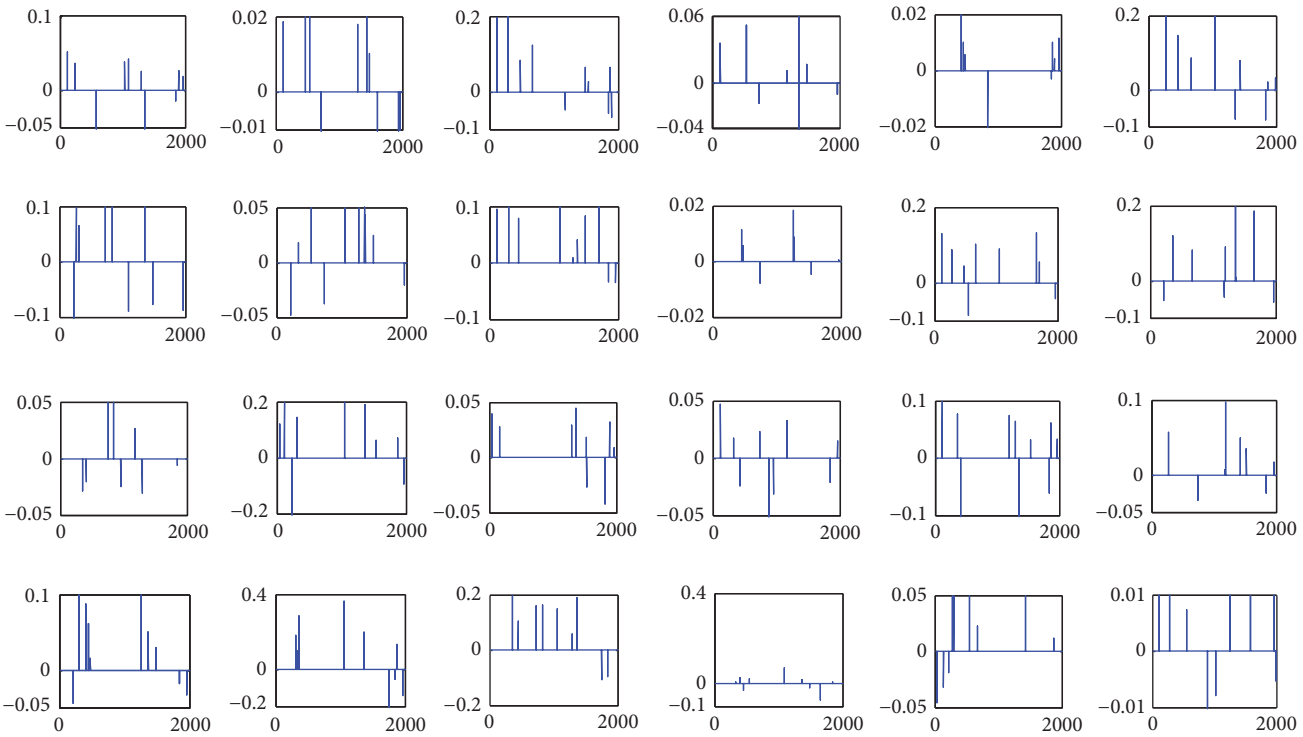


Figure 8: Sparse coefficients N of the SR with 24 total channels.

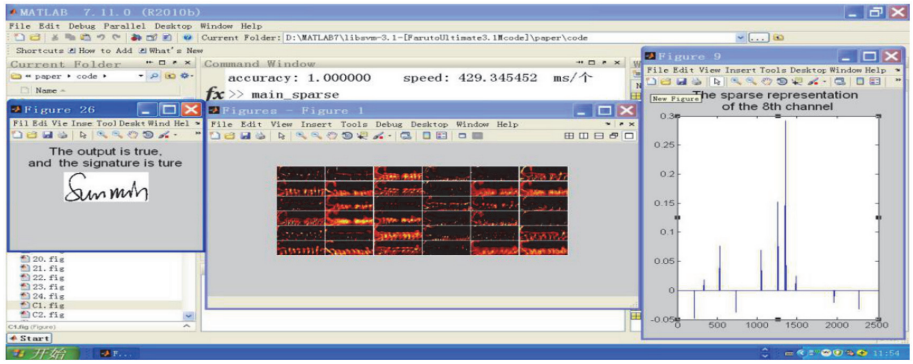


Figure 9: Signature authentication process and results.

on balancing and/or improving the performance between the signature authentication accuracy and the computation time of the proposed SIA algorithm.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was sponsored by the Natural Science Foundation of China (61300179), Key Scientific and Technological Project of Shaanxi Province (2016GY-040), and the Science Foundation of Xi'an University of Science and Technology (104-6319900001). We also thank Master students Huan Li and Min Sun for their support of data set collection.

References

[1] H. Lu, Y. Li, Y. Zhang, M. Chen, S. Serikawa, and H. Kim, "Underwater Optical Image Processing: a Comprehensive Review," *Mobile Networks & Applications*, vol. 2017, pp. 1–12, 2017.

[2] A. Singh, P. Singh, A. Amini, and F. Marvasti, "Twin tree hierarchy: a regularized approach to construction of signature matrices for overloaded CDMA," *Wireless Communications and Mobile Computing*, vol. 16, no. 17, pp. 3070–3088, 2016.

[3] L. Yuan, Q. Ran, and T. Zhao, "Image authentication based on double-image encryption and partial phase decryption in nonseparable fractional Fourier domain," *Optics & Laser Technology*, vol. 88, pp. 111–120, 2017.

[4] H. Wang, Y. Qin, Y. Huang, Z. Wang, and Y. Zhang, "Multiple-image encryption and authentication in interference-based scheme by aid of space multiplexing," *Optics & Laser Technology*, vol. 95, pp. 63–71.

[5] K. S. Manjunatha, S. Manjunath, D. S. Guru, and M. T. Somashekara, "Online signature verification based on writer dependent features and classifiers," *Pattern Recognition Letters*, vol. 80, pp. 129–136, 2016.

[6] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 41–50, 2014.

[7] S. Lai, L. Jin, and W. Yang, "Toward high-performance online HCCR: A CNN approach with DropDistortion, path signature and spatial stochastic max-pooling," *Pattern Recognition Letters*, vol. 89, pp. 60–66, 2017.

[8] E. N. Zois, L. Alewijnse, and G. Economou, "Offline signature verification and quality characterization using poset-oriented grid features," *Pattern Recognition*, vol. 54, pp. 162–177, 2016.

[9] A. Sharma and S. Sundaram, "An enhanced contextual DTW based system for online signature verification using Vector Quantization," *Pattern Recognition Letters*, vol. 84, pp. 22–28, 2016.

[10] Y. Fang, W. Kang, Q. Wu, and L. Tang, "A novel video-based system for in-air signature verification," *Computers & Electrical Engineering*, vol. 57, pp. 1–14, 2017.

[11] M. Wang, W.-Y. Chen, and X. D. Li, "Hand gesture recognition using valley circle feature and Hu's moments technique for robot movement control," *Measurement*, vol. 94, pp. 734–744, 2016.

[12] B. Kovari and H. Charaf, "A study on the consistency and significance of local features in off-line signature verification," *Pattern Recognition Letters*, vol. 34, no. 3, pp. 247–255, 2013.

[13] J. Canuto, B. Dorizzi, J. Montalvão, and L. Matos, "On the infinite clipping of handwritten signatures," *Pattern Recognition Letters*, vol. 79, pp. 38–43, 2016.

[14] R. Plamondon, C. O'Reilly, J. Galbally, A. Almaksour, and É. Anquetil, "Recent developments in the study of rapid human movements with the kinematic theory: Applications to handwriting and signature synthesis," *Pattern Recognition Letters*, vol. 35, no. 1, pp. 225–235, 2014.

[15] A. Soleimani, B. N. Araabi, and K. Fouladi, "Deep Multitask Metric Learning for Offline Signature Verification," *Pattern Recognition Letters*, vol. 80, pp. 84–90, 2016.

[16] A. B. De Oliveira, P. R. Da Silva, and D. A. C. Barone, "A novel 2D shape signature method based on complex network spectrum," *Pattern Recognition Letters*, vol. 63, article no. 6248, pp. 43–49, 2015.

[17] M. Wang, L. Guo, and W.-Y. Chen, "Blink detection using Adaboost and contour circle for fatigue recognition," *Computers and Electrical Engineering*, vol. 58, pp. 502–512, 2017.

[18] M. Wang, W. Qu, and W.-Y. Chen, "Hybrid sensing and encoding using pad phone for home robot control," *Multimedia Tools and Applications*, pp. 1–14, 2017.

[19] M. Wang, S. Zhang, Y. Lv, and H. Lu, "Anxiety Level Detection Using BCI of Miner's Smart Helmet," *Mobile Networks and Applications*, pp. 1–8, 2017.

[20] Y. Liu, Z. Yang, and L. Yang, "Online Signature Verification Based on DCT and Sparse Representation," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2498–2511, 2015.

[21] P. Porwik, R. Doroz, and T. Orczyk, "Signatures verification based on PNN classifier optimised by PSO algorithm," *Pattern Recognition*, vol. 60, pp. 998–1014, 2016.

[22] H. Lu, Y. Li, T. Uemura et al., "FDCNet: filtering deep convolutional network for marine organism classification," *Multimedia Tools and Applications*, pp. 1–14, 2017.

[23] N. Kaothanthong, J. Chun, and T. Tokuyama, "Distance interior ratio: A new shape signature for 2D shape retrieval," *Pattern Recognition Letters*, vol. 78, pp. 14–21, 2016.

[24] X. Xia, Z. Chen, F. Luan, and X. Song, "Signature alignment based on GMM for on-line signature verification," *Pattern Recognition*, vol. 65, pp. 188–196, 2017.

[25] S. Ma, G. Chen, W. Wu, L. Song, X. Tian, and X. Wang, "Identifying effective initiators in OSNs: from the spectral radius perspective," *Wireless Communications and Mobile Computing*, vol. 16, no. 18, pp. 3340–3359, 2016.

[26] G. Goswami, P. Mittal, A. Majumdar, M. Vatsa, and R. Singh, "Group sparse representation based classification for multi-feature multimodal biometrics," *Information Fusion*, vol. 32, no. 1, pp. 3–12, 2016.

[27] H. Liu, S. Li, and L. Fang, "Robust Object Tracking Based on Principal Component Analysis and Local Sparse Representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 11, pp. 2863–2875, 2015.

[28] B. Hou, B. Ren, G. Ju, H. Li, L. Jiao, and J. Zhao, "SAR image classification via hierarchical sparse representation and multisize patch features," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 33–37, 2016.

[29] Y. Shao, N. Sang, C. Gao, and L. Ma, "Probabilistic class structure regularized sparse representation graph for semi-supervised hyperspectral image classification," *Pattern Recognition*, vol. 63, pp. 102–114, 2017.

[30] J. H. Cheon and M.-K. Lee, "Improved batch verification of signatures using generalized sparse exponents," *Computer Standards & Interfaces*, vol. 40, pp. 42–52, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 3132424, 9 pages
https://doi.org/10.1155/2018/3132424

Research Article

Cognitive-Empowered Femtocells: An Intelligent Paradigm for Femtocell Networks

Xiaoyu Wang,[1] Pin-Han Ho,[1] Alexander Wong,[1] and Limei Peng[2]

[1]*Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada N2L 3G1*
[2]*School of Computer Science and Engineering, Kyungpook National University, Republic of Korea*

Correspondence should be addressed to Limei Peng; auroraplm@knu.ac.kr

Received 29 March 2018; Accepted 18 April 2018; Published 20 June 2018

Academic Editor: Haider Abbas

Deploying femtocells has been taken as an effective solution for removing coverage holes and improving wireless service performance in 3G-beyond wireless networks such as WiMAX and Long Term Evolution (LTE). This article investigates a novel framework of dynamic spectrum management for femtocell networks, called cognitive-empowered femtocells (CEF), aiming at mitigating both cross-tier and intratier interferences with minimum modifications required on the corresponding macrocell network. With the proposed framework, each CEF base station (BS) and the femtocell users can utilize spatiotemporally available radio resources for the access traffic. We conclude that the proposed CEF framework can effectively complement the existing femtocell design and serve as a value-added feature to the state-of-the-art femtocell technologies, while achieving high scalability and interoperability by minimizing the required modifications on the macrocell protocol design.

1. Introduction

With a wide deployment of 3G and/or 3G-beyond wireless network infrastructure, ubiquitous communications and heterogeneous service provisioning can be achieved by an integrated macrocellular networking strategy that supports all the users in a single stage no matter they are mobile or fixed, indoor or outdoor, and for data or voice services [1]. A lesson learned from early experiences in developing macrocellular networks is that it is expensive to support both line-of-sight and non-line-of-sight communications in a typical range of a few tens of kilometers, and it becomes less economically viable to build infrastructures with increasing data rates. Besides, quality of service (QoS) could be noticeably degraded by the path loss, shadowing, and multipath fading effects due to wall penetration, which result in low data rates and poor voice quality inside the buildings.

One of the recent advances for overcoming indoor communication barriers without taking much infrastructure expenditure is the use of femtocells, which can achieve high data rate and manageable QoS for both users of macrocell and indoor femtocells. A femtocell is a small cellular area covering homes or offices, while a femtocell base station (BS) is simple, low-cost, and miniature access point designed for indoor wireless service coverage of the corresponding macrocell. As such, femtocell networks are end-user deployed hotspots that underlay the planned macrocell networks of mobile operators. Instead of using wireless transmissions like relays, a femtocell BS is connected with the macrocell BS via wired lines, such as coaxial cables. A femtocell provisions services as a whole with the corresponding macrocell BS, where a femtocell user could consume the femtocell resources at this moment yet switch to the macrocell and become a macrocell user in the next moment due to mobility, possibly by using different sets of radio resources and vice versus. Thus, a two-tier architecture is formed with the macrocell BS in the first/top tier and the femtocell BSs in the second/below tier.

1.1. Interference Resolution in Femtocell Networks. According to the most recent development in the 3GPP LTE/LTE-Advanced standardization progress, femtocells have the following three deployment modes [2–4]:

(i) Dedicated channel deployment: the femtocell and the macrocell utilize radio spectrum orthogonal to each other.

(ii) Cochannel deployment: the femtocell and the macrocell utilize a common set of radio subbands.

(iii) Partial cochannel deployment: some parts of radio spectrum utilized by the femtocell are orthogonal to that of the macrocell, while other parts of radio spectrum utilized by the femtocell overlap with that of the macrocell.

Although the dedicated channel deployment can avoid cross-tier interferences, the limited bandwidth of both femtocells and macrocell could seriously impair the performance. It is particularly not feasible under dense deployment of femtocells since each femtocell can only access very limited bandwidth. In the cochannel and partial cochannel deployment, on the other hand, a global scheduling scheme is needed for channel allocation; otherwise both the femtocells and the macrocell may suffer from terrible interference with each other. This becomes a major challenge in adopting these schemes.

A number of interference management approaches for femtocell networks have been reported, including a power control strategy for femtocell users [5], time hopped CDMA (TH-CDMA) combined with sectorized antenna [6], signal-to-interference-plus-noise based component carrier selection [7], and a centralized scheduling scheme that considers the mutual interference of both femtocell and macrocell users [8]. However, due to the design requirement for simplicity with minimum modifications on the macrocell protocols running at the BS, these interference management approaches may not be efficient and scalable. Note that the coverage of a macrocell could be over thousands of femtocells. Therefore, it is not a scalable solution in jointly considering those femtocell users in the design of macrocell resource allocation and scheduling schemes.

1.2. Does Cognitive Radio Take a Role? The concept of Cognitive Radio (CR) [9] was introduced decades ago, and its goal is to utilize spatiotemporally unused spectrum resources of licensed radio spectrums in a secondary and opportunistic manner, without interfering with the licensed user signals. Therefore, the CR techniques have been considered as attractive solutions to the improvement of spectrum utilization and mitigation of spectrum resource starvation for ubiquitous wireless services. The first standard aimed at using the CR techniques is IEEE 802.22, of which the initial drafts specify that the CR enabling networks should operate in spectrum allocated to Ultra High Frequency/Very High Frequency (UHF/VHF) TV broadcast service with a Point-to-Multipoint (P2MP) centralized infrastructure [3, 4].

To solve the interference management problem without imposing additional complexity and deviation from the current macrocell network design, we turn to consider the approach of CR dynamic spectrum management, which has demonstrated a strong synergy with the femtocell interference management in terms of their design premises and principle missions.

1.3. Outline of CEF Framework. This article investigates a novel framework of cognitive-empowered femtocells (CEF) for achieving efficient management for both cross-tier and intratier interferences. The CR dynamic spectrum sensing technique is taken as a built-in feature of the CEF BSs and the femtocell user handsets and aims to serve as an effective complement to the existing femtocell technologies. Under the CEF framework, the radio resources that a femtocell user can use to communicate with the CEF BSs are not only the licensed spectrum of the macrocell, but also the spectrum allocated to UHF/VHF TV broadcast services. This is expected to achieve effective interference management by minimizing the cross-tier and intratier interference via an opportunistic manner with little modification required on the macrocell protocol design. Besides, the consideration on other licensed bands can further resolve the possible bandwidth thirsty and improve QoS in femtocell networking.

To effectively and dynamically identify spatiotemporally available spectrum under the CEF framework, we propose a novel sensing coordination scheme for initiating interference-free communications between a CEF BS and its femtocell users. We will show that the proposed scheme can perfectly fit to the unique features and design premises of femtocell networks while taking the best advantage of the conventional standalone sensing and cooperative sensing strategies [10].

2. Dynamic Spectrum Sensing under CEF

Dynamic spectrum sensing is a unique feature in CR networks, which concerns whether an efficient and interference-free spectrum reuse at a CR device can be achieved. This section provides an overview on the state-of-the-art spectrum sensing technologies and the recently reported dynamic spectrum sensing schemes.

2.1. Existing Spectrum Sensing Techniques. There have been many spectrum sensing techniques proposed for radio-scene analysis [11], such as energy detection, cyclostationary detection, pilot-based coherent detection, and covariance-based detection. Due to the low computational complexity and easy implementation, energy detection is a natural choice for wideband sensing in CR networks [12]. The underlying motivation for using wideband sensing for CR networks is the desire to obtain as much vacant spectrum resources as possible in a simultaneous manner. Much of the focus has been placed in wideband sensing techniques on the signal reconstruction and detection decision making process [13–15].

An approach for reliable power spectral density (PSD) estimation and subband identification serves as a basis in achieving efficient wideband sensing when dealing with the heterogeneity of subbands over the spectrum. This is essential to the proposed CEF framework that deals with a dynamic network environment with changing subband availabilities and noise fluctuation conditions, as well as a heterogeneous mix of different macrocell users with different spectrum resource requirements. In this article, a state-of-the-art wideband sensing technique in [16] is employed for this purpose. It is characterized by a strong ability of noise fluctuation-free PSD estimation for wideband sensing as well as automatic detection of subband information from the

acquired radio frequency signal, which is considered perfectly for the CEF BSs design in the energy detection based proactive sensing process under the proposed framework.

2.2. Dynamic Spectrum Sensing Schemes. A spectrum sensing technique takes a suitable dynamic sensing scheme as the implementation approach in a particular system. Cooperative sensing and standalone sensing are two most popular dynamic spectrum sensing schemes that have been widely reported and employed. With cooperative sensing, a central controller (e.g., a base station, cluster head, or a data sink) exercises a complete control over when and which subbands each wireless node must sense and makes a final decision on the spectrum availability by fusing the sensing results obtained from all the wireless nodes. With standalone sensing, on the other hand, each secondary user individually controls over when and which subbands to sense, and spectrum availabilities are solely determined based on its own observations. The advantages and disadvantages associated with both strategies are provided in Table 1.

Clearly, both sensing schemes are subject to limitations on the suitability for realizing the proposed CEF framework. For example, due to the nonnegligible heterogeneity in modern wireless networks which may accommodate devices of different vendors and wireless techniques, using a central controller to handle spectrum sensing, access and resource allocation processes may result in problems of low flexibility, poor scalability, and bad interoperability. In addition, the CEF BSs and femtocell users are simple devices, and it could be infeasible to implement a complicated distributed computing platform for cooperative sensing. Conversely, given that all the spectrum sensing and information processing are performed independently without any additional information and coordination, resolving the aforementioned problems could lead to significant performance degradation due to disorder accesses by secondary users, which then affect the sensing precision.

2.3. Sensing Coordination. The proposed CEF framework implements a coordinated spectrum sensing scheme, aiming to take the best of the two conventional spectrum sensing schemes, say cooperative sensing and standalone sensing, while avoiding the respective disadvantages, in order to satisfy the specific design requirements of the CEF networks. With coordinated sensing, when more radio resources are needed than that is available from the macrocell licensed bands, standalone sensing is performed under the coordination of associated CEF BS, which schedules the spectrum sensing process of all the surrounding femtocell users. Since the CEF BS simply instructs the sensing sequence and the range of spectrum for sensing instead of concluding the spectrum availability for each femtocell user, the master/slave relation between CEF BS and femtocell users is loosened when compared with cooperative sensing.

It is expected that integrating the sensing coordination scheme on top of the existing femtocell technologies allows for greater spectrum resource usage and efficient interference management, hence acting as a value-added complement to the femtocell network design. Its simplicity and efficiency, along with the high transparency to the existing protocols in the system, perfectly satisfies the desired features and original premises of modern femtocell systems.

3. The Proposed CEF Framework

The section first defines the functional modules of the CEF BSs, with a particular focus on the ones for dynamic spectrum sensing purposes, followed by the description of the proposed coordinated sensing scheme.

3.1. Functions of CEF BSs. The most distinguished feature of the CEF network is that a CEF BS can initiate communications with its femtocell users via both macrocell bands and unused TV bands in an opportunistic manner, so as to enlarge the system capacity while intelligently mitigating both cross-tier and intratier interference. Thus, in addition to the functions of existing femtocells, a CEF BS and its femtocell users explore both the macrocell's licensed bands as well as the licensed spectrum resources for UHF/VHF TV by periodically sensing the resource blocks via energy detection. The sensing at the CEF BS is to maintain real-time environmental and channel information related to the femtocell users. With the preliminary sensing results, the CEF BS schedules the femtocell users on when and which subbands to sense in a stochastic manner. It is clear that disorderly sensing among a group of femtocell users on a common set of available subbands could lead to incorrect sensing results and cause undesired interference [10]. Thus, under the coordination of the CEF BS, the femtocell users are instructed in terms of the range and priority of the possible channels to scan, and the coordination is expected to significantly improve the sensing accuracy at each femtocell user. Compared with cooperative sensing, the proposed coordinated sensing scheme has each femtocell user scan over a much smaller set of subbands with some certain sequence, which can effectively reduce the sensing delay and power consumption.

The CEF function is composed of two functional modules: *sensing coordination module* (SCM) installed at the CEF BS and *end-user modules* (EUM) equipped in the femtocell user handsets.

The operation at each SCM is comprised of three main processing phases: (i) proactive sensing, (ii) sensing coordination, and (iii) acknowledge (ACK) information adjustment. The operation of each EUM is comprised of two main processing phases: (i) knowledge-based estimation and (ii) sensing under reasoning.

The relations among the functional modules in SCM and EUM are shown in Figure 1. In the proactive sensing phase, the CEF BS periodically performs channel measurement to identify available resource blocks and collect real-time and immediate channel information. In each measurement period, the CEF BS performs sensing coordination by stochastically selecting likely available subbands and accordingly sharing the information with the EUM. This is to avoid cross-tier interferences.

With the shared information, the EUM estimate the sensing parameters such as the maximum number of sensing iterations and channel sequence for sensing and perform fine

Table 1: Comparison between cooperative and stand-alone sensing.

	Cooperative Sensing	**Stand-alone Sensing**
Density	Highly dependent on the density of sensing nodes. Highly dependent on independency of observation.	N/A
Heterogeneity Signals	May result in the situation where secondary users submit different sensing results and conclusions due to different perceptions of heterogeneity signals.	N/A
Communication Overhead	Heavier communication overhead.	N/A
Time Sensitivity	Introduce additional delay from collecting sensing results from sensing nodes to making the decision, resulting staled sensing results.	Able to promptly use available sub-bands once identifying the availabilities.
Spatial Diversity Gain	Achieve spatial diversity gain with certain densities and uncorrelated observations.	Not able to achieve.
Fading Signals	Higher detection sensitivity under proper density.	Able to achieve detection sensitivity by using feature detection techniques, such as cyclostationary detection, and covariance-based detection.
Coordination	Sensing and access are fully controlled by the central controller.	May result in disorderly accesses to the same available sub-bands like a swarm of bees.
Reliability	Highly depends on the data fusion scheme as well as the credibility of sensing nodes.	Highly depends on sensing techniques.

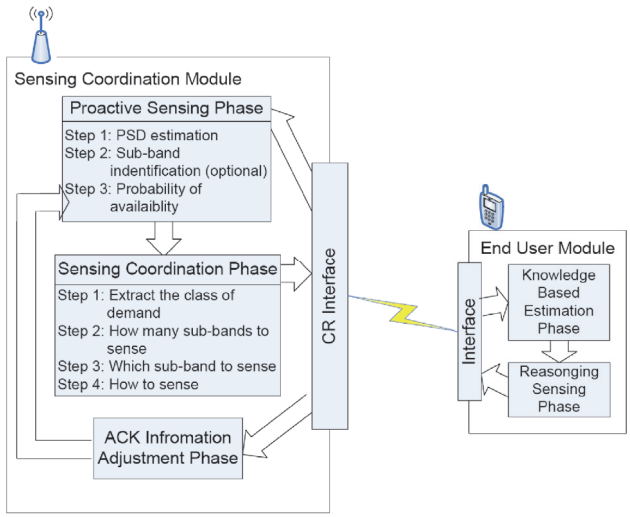


Figure 1: An overview of sensing coordination in CEF framework. It consists of the sensing coordination module (SCM) module on the left and the end-user module (EUM) on the right. SCM consists of two phases, i.e., proactive sensing phase and sensing coordination phase. EUM consists of two phases, i.e., knowledge-based estimation phase and reasoning sensing phase.

sensing on the suggested subbands via a reasoning process, in order to minimize the intratier interferences. Once a femtocell user accesses the subbands, the femtocell user will acknowledge the CEF BS regarding the usage of the subbands. Such ACK information assists the SCM to perform channel measurement on the corresponding spectrum to identify the macrocell user signals. Note that if any macrocell user signal is identified, the SCM will instruct the femtocell users to evacuate from those subbands immediately.

3.2. Sensing Coordination Modules (SCM). In brief, the SCM proactively performs wideband sensing for channel measurement to identify transmitting/receiving opportunities beyond legacy technologies, as well as maintaining and updating the proactive sensing results, so as to assist the femtocell users to determine their respective sensing parameters when they perform standalone spectrum sensing. The sensing parameters that will be estimated at each femtocell user handset will be introduced later.

In the following discussions, we consider the time and frequency under OFDMA/TDMA as network resources and use them as *resource blocks*. The periodic channel measurement process is illustrated in Figure 2 and summarized as follows:

(i) Proactive sensing phase: the SCM measures the Received Interference Power (RIP) on each resource block (i.e., each time slot of each subband). A measurement is performed during a complete subframe via wideband sensing upon all the time slots and subbands in the subframe, as shown in Figure 2.

(ii) Sensing coordination phase: the SCM extracts site-information of the macrocell from the wideband sensing and shares the information with the femtocell users.

(iii) ACK information adjustment phase: the CEF BS can adjust the original detection threshold to better estimate the activity of the associated macrocell users according to the ACKs information from its femtocell users.

Note that a resource block is considered as occupied by the macrocell if the RIP on that resource block exceeds a certain threshold. In any subframe not performing measurement, the CEF BS only schedules unoccupied resource blocks by the macrocell to its users. The SCM also obtains the channel availability information possibly on other licensed bands, such as that for TV.

The above three phases are performed in each channel measurement period of the SCM, which are detailed in the following subsections.

3.2.1. Proactive Sensing Phase

(i) **Step 1.** A fluctuation-free power spectral density (PSD) is obtained by using a constrained Bayesian estimation approach [16], where a weighted average of each frequency is computed across a tapering window. The weight assigned to each frequency within the tapering window is set based on a Gibbs-based likelihood function, where a low weight is assigned when the PSD deviation is high between that frequency and the frequency we wish to estimate. This likelihood function serves for two purposes: (i) frequencies with low PSD deviations (where the PSD deviations are due largely to noise fluctuations) have high contribution in smoothing out the noise fluctuations and (ii) frequencies with high PSD deviations (where the PSD deviations are due largely to the underlying PSD shape) have low contribution and as such have little effect on the overall PSD shape. Therefore, noise fluctuations are suppressed while the overall PSD shape is preserved.

(ii) **Step 2.** In the situation where the subbands are unknown (e.g., wireless devices from different manufacturers use different frequency bands), a first-order derivative filter is applied to the fluctuation-free PSD to identify significant PSD changes, which characterizes the boundaries of the individual subbands [16].

(iii) **Step 3.** Test statistics are computed from the fluctuation-free PSD within each subband and compared it with the detection threshold to determine the probability of availability. Furthermore, the threshold will be updated in the ACK information adjustment phase.

3.2.2. Sensing Coordination Phase. In the sensing coordination phase, the SCM instructs the femtocell users on which subbands and in what sequence to sense these subbands, in order to mitigate intratier interferences and improve the likelihood of obtaining usable bandwidth and meeting the tolerable delay due to the sensing process. On the other hand, it leaves the decision on transmission rate and modulation to

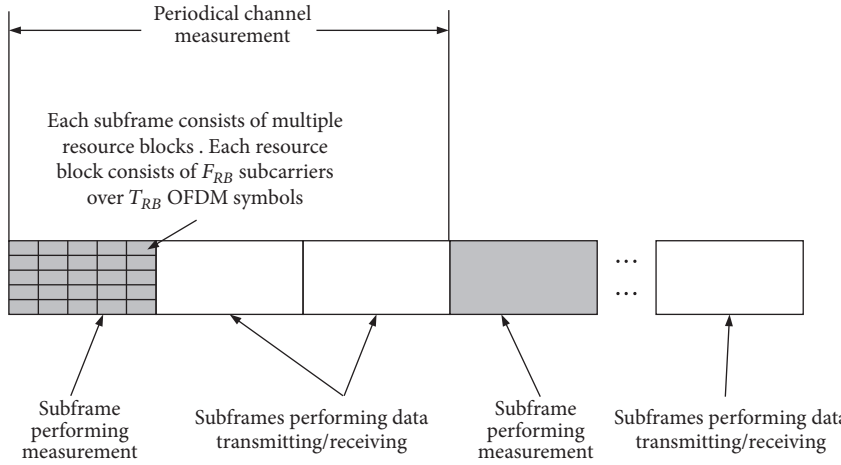


Figure 2: The femtocell BS performs periodical channel measurement to identify whether a resource block is occupied by the macrocell in a subframe.

the femtocell users, which are essential conditions to satisfy their QoS requirements.

Note that there could be a hidden terminal problem at the femtocell user side when a subband that is considered available at the CEF BS side is actually occupied by an adjacent CEF BS. This simply causes intratier interferences. Therefore, instead of completely following the instruction of the channel availability from the CEF BS, each user handset performs fine sensing via energy detection on the set of subbands instructed by the CEF BS through a reasoning approach.

(i) **Step 1.** Extract the service class c_m of demand.

(ii) **Step 2.** Stochastically determine the number of subbands, denoted as N_{c_m}, for the femtocell users to sense with a probability θ_{c_m} according to a distribution $f_{c_m}(\overline{N}_{c_m}, \sigma_{c_m})$ with mean $\overline{N}_{c_m}$ and standard deviation σ_{c_m}. Note that this can be a simple Gaussian distribution.

(iii) **Step 3.** Stochastically select a set of N_{c_m} subbands for the femtocell users to sense and access according to the probability of subband availability determined in the proactive sensing phase. As such, the femtocell users are instructed on the most likely available subbands.

(iv) **Step 4.** Among the set of selected subbands, instruct the femtocell users to perform only energy detection prior to access on the subbands associated with the sensing results within the channel detection time, denoted as τ, which is specified in IEEE 802.22. Energy detection is sufficient for these cases due to the freshness of the sensing results. If the recommended subbands are older than τ seconds, the femtocell users will be instructed to perform feature detection prior to access given the staleness of the sensing results.

Fairness is another important design goal other than interference mitigation, which can be effectively achieved by manipulating the number of subbands that a user is allowed to scan for each transmission. On the other hand, the optimal number of subbands for scanning at a femmtocell user can also help to mitigate intratier interference since the femtocell users, with a stochastic channel sensing strategy, do not likely select a common set of subbands and scan them in the same order. This will be examined in the simulations.

3.2.3. ACK Information Adjustment Phase. In the ACK information adjustment phase, the ACK messages of the femtocell users bear the information on which subbands have been used by the femtocell users and the statistics of the channel conditions, such that the CEF BS can adjust the original detection threshold to better estimate the activity of the associated macrocell users.

3.3. End-User Modules (EUM). The EUM at femtocell user handsets take advantage of both extrinsic and intrinsic knowledge of the network environment to estimate the optimal number of channels in their standalone sensing process, which can use either energy detection or feature detection, according to the instructions from the SCM at the CEF BS. Moreover, the EUM dynamically refine the amount of sensing results to achieve the desired QoS requirements. The two phases are further elaborated in the following subsections.

3.3.1. Knowledge-Based Estimation Phase. Intuitively, scanning more channels will more likely obtain sufficient bandwidth to support the desired connections if sensing time is not a concern. However, sensing more subbands results in longer sensing delay that consequently decreases the throughput. Besides, a long sensing process increases the likelihood of unsuccessful transmissions and lost opportunities due to the dynamic nature of channel availability. Therefore, the cognitive EUM estimate the number of subbands to scan (denoted as n^*) using the knowledge instructed by the CEF BS, in order to achieve the best customer premise.

3.3.2. Reasoning Sensing Phase. Since EUM has to scan subbands one after the other, it is possible that an EUM

identifies sufficient channels before scanning all the subbands recommended by the CEF BS. Therefore, a reasoning process is suggested to improve the sensing efficiency, where each EUM decides to proceed to scan the next subband only if the expected return in terms of throughput is positive, and the number of subbands has not reached n^*.

4. Performance Evaluation

Experiments were conducted to verify the performance of the proposed CEF framework. We simulated $50m \times 50m$ indoor network area with an average of 10 macrocell users and various number of femtocell users, which were allocated according to a Pareto distribution in three femtocells. Each femtocell user has a radio transmission range radius of $R = 30$. For each spectrum sensing event, a $400MHz$ spectrum is divided into 10 subbands with randomly generated bandwidth (which reflects the stochastic nature of wireless channels). For each transmission, a femtocell user is randomly chosen and then the intended receiver is the nearest CEF BS. We conducted the simulation for $t_{sim} = 5000s$ for each trial, where the performance of the proposed CEF framework was evaluated in terms of (1) the throughput upper bound in the proactive sensing phase, (2) the probability of intratier interference, and (3) the temporal usage rate. Note that we assume the wideband sensing in (1) can accurately identify the available resource blocks, such that the cross-tier interference can be completely removed.

4.1. Performance of Proactive Sensing via Periodic Channel Measurement. It is clear that the available channels provided by the CEF BS may or may not be free at the femtocell users, depending on the fine sensing results at the femtocell users which clarify the intratier interferences. Therefore, the throughput achieved by using available resource blocks identified in the proactive sensing phase can be simply taken as an upper bound on the real achievable throughput after considering the sensing results at the femtocell user side. This is illustrated in Figure 3, where the proposed scheme effectively achieves significant enhancement when compared with the scenario of "no measurement". Note that with "no measurement" the CEF BS randomizes the resource block allocation in each subframe, which is similar to the concept of interleaved resource block allocation to combat the block fading channel. The measurement period is counted by subframes.

4.2. Probability of Intratier Interference. This set of simulations evaluates the probability of intratier interference, which is the probability of more than one femtocell user identifying any common available subband. This causes intratier interference since competition for medium access could arise hereafter.

The probability of intratier interference is a direct performance measure on the proposed framework, and it determines the feasibility and effectiveness of the proposed framework. In Figure 4, the probability of intratier interference in the proposed CEF framework, the standalone sensing scheme in [17], and the cooperative scheme is compared with the

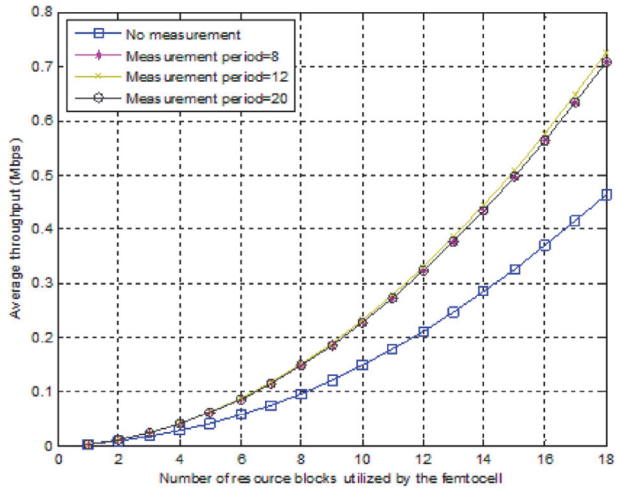


Figure 3: Simulation results of the average throughput of the femtocell.

number of subbands being set to 5. As expected, the proposed framework yields a much lower probability of intratier interference than that by the standalone scheme and comparable with that of the cooperative scheme. The result demonstrates the potential benefits of integrating coordinated sensing into the CEF framework, where a spectrum sensing process with similar simplicity of standalone sensing yet the same efficiency as cooperative sensing can be achieved. Further, with more femtocell users, the proposed framework can better outperform than the standalone scheme in terms of the probability of intratier interference.

We noticed that with cooperative sensing the probability of intratier interference slightly decreases as the number of femtocell users increases. It is because the sensing accuracy of the proposed framework is highly dependent on the density of femtocell users and remains stable after reaching a certain number of femtocell users [18]. The results demonstrate the efficiency that can be achieved through the use of the proposed framework.

4.3. Temporal Usage Rate. To further evaluate the network-wide temporal efficiency, we investigated the temporal usage rate, which is defined as the percentage of time that an arbitrary subband is not used by any femtocell user. Our goal is to evaluate the impact of the femtocell user traffic on the underlay network-wide performance. By setting the communication traffic volume of each macrocell user as 10 packets/second, Figure 5 shows the temporal usage rate in the proposed framework with different femtocell user traffic volumes. It can be observed that the temporal usage rate is noticeably higher than that of the standalone sensing scheme due to the fact that the CEF BSs instruct the most likely available subbands for femtocell users based on both *a priori* subband information obtained from its energy detection and ACK information adjustment. Moreover, the proposed framework can achieve a similar level of network-wide temporal efficiency when compared with that of the conventional cooperative sensing scheme, while significantly

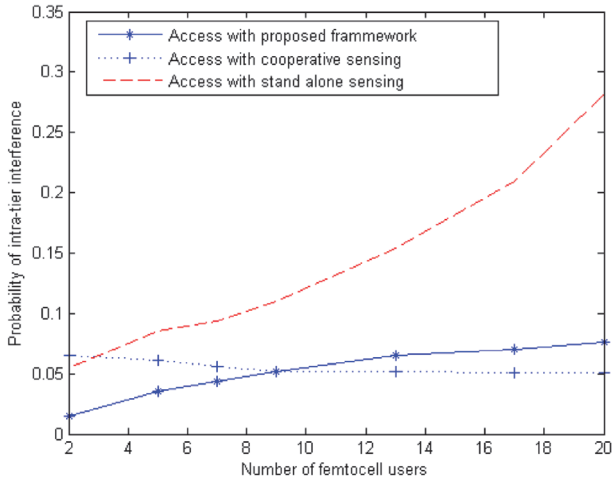


FIGURE 4: Probability of intratier interference versus the number of femtocell users, $N_{(2)}$.

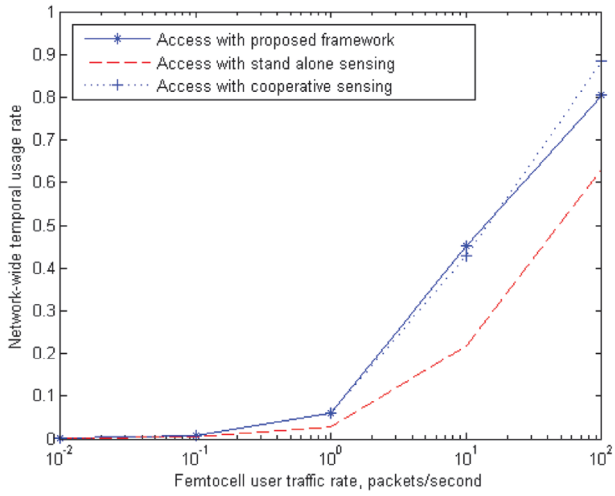


FIGURE 5: Temporal usage rate of the proposed framework with different femtocell user traffic arrival rate.

outperforming the cooperative sensing scheme in terms of overall efficiency and simplicity. This is a promising feature of the proposed framework with the effort of deploying simple, low cost, and custom-premised CEF BSs.

5. Conclusions

In this article, we proposed a novel framework of interference management by way of channel measurement and dynamic spectrum sensing for femtocell networks, called cognitive-empowered femtocells (CEF), aiming at enhancing the femtocell capacity and mitigating both cross-tier and intratier interference in a single step. Under the proposed framework, the CEF BSs periodically perform channel measurement and sensing coordination with the corresponding femtocell users. With the dynamic spectrum sensing capabilities, the devices can utilize spatiotemporally available spectrum in an opportunistic manner. Simulation results demonstrated the potential of the proposed CEF framework with the effort of improving the overall network capacity via intelligent acquisition of spectrum opportunities with excellent scalability, flexibility, and transparency to the existing macrocell protocol design.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (2015R1C1A1A02036536).

References

[1] "The Evolution of Mobile Technologies," in *1G 2G 3G 4G LTE, Qualcomm*, The Evolution of Mobile Technologies, 1G 2G 3G 4G LTE, 2014.

[2] 3GPP R4-091976, "LTE-FDD HeNB interference scenarios," in *RAN4 #51*, 2009.

[3] X. Y. Wang, P.-H. Ho, and K.-C. Chen, "Interference analysis and mitigation for cognitive-empowered femtocells through stochastic dual control," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2065–2075, 2012.

[4] K. A. Meerja, P.-H. Ho, B. Wu, and H.-F. Yu, "A novel architecture and media access protocol for cognitive radio based autonomous femtocell networks," in *Proceedings of the 10th IEEE and IFIP International Conference on Wireless and Optical Communications Networks, WOCN 2013*, ind, July 2013.

[5] C. Tseng et al., "Itigating Uplink Interference in Femto-Macro Coexisted Heterogeneous Network by Using Power Control," *Journal of Wireless Personal Communications*, vol. 95, no. 1, 2017.

[6] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, 2008.

[7] Y. Chen, J. Zhang, and Q. Zhang, "Utility-aware refunding framework for hybrid access femtocell network," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1688–1697, 2012.

[8] C. Bouras, "Interference Management Strategy for 5G Femtocell Cluster," *Wireless Personal Communications*, vol. 96, no. 1, 2017.

[9] M. Amjad et al., "Full-Duplex Communication in Cognitive Radio Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, 2017.

[10] X. Y. Wang and P.-H. Ho, "A Novel Sensing Coordination Framework for CR-VANETs," *IEEE Transactions on Vehicular Technology Special Issue on Cognitive Radio*, 2009.

[11] T. Yucek et al., "A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, 2009.

[12] S. Bae, "On Optimal Cooperative Sensing with Energy Detection in Cognitive Radio," *MDPI Sensors*, vol. 17, no. 9, 2017.

[13] 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.

[14] X. Liu et al., "5G-based wideband cognitive radio system design with cooperative spectrum sensing," *Physical Communication*, vol. 25, no. 2, 2017.

[15] J. Font-Segura et al., *Wideband Cognitive Radio: Monitoring, Detection and Sparse Noise Subspace Commuinication [Ph.D. thesis]*, Technical University of Catalonia, July 2014.

[16] A. Wong, "Constrained Bayesian Power Spectral Density Estimation and Sub-band Identification for Wideband Spectrum Sensing in Cognitive Radio," http://www.einfodaily.com/wbss/wbss_submitted.pdf.

[17] X. Y. Wang, A. Wong, and P.-H. Ho, "Extended knowledge-based reasoning approach to spectrum sensing for cognitive radio," *IEEE Transactions on Mobile Computing*, vol. 9, no. 4, pp. 465–478, 2010.

[18] J. A. del Peral-Rosado, M. Bavaro, J. A. Lopez-Salcedo et al., "Floor Detection with Indoor Vertical Positioning in LTE Femtocell Networks," in *Proceedings of the 2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, San Diego, CA, USA, December 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6183732, 11 pages
https://doi.org/10.1155/2018/6183732

Research Article

Framework for E-Health Systems in IoT-Based Environments

Maruf Pasha and Syed Muhammad Waqas Shah

Department of Information Technology, Bahauddin Zakariya University, Multan, Pakistan

Correspondence should be addressed to Maruf Pasha; maruf.pasha@bzu.edu.pk

Received 30 October 2017; Revised 14 January 2018; Accepted 4 February 2018; Published 7 June 2018

Academic Editor: Yin Zhang

Internet of things technology consists of physical objects that are accessible via the Internet, for example, devices, vehicles, and buildings. Internet of things technology is used to connect these physical objects by utilizing the existing infrastructure of networks. A unique identifier is assigned to identify the objects in IoT environments. Internet of things technology is used to make productive decisions on the sensed data after converting it into information. IoT technology is being used in various life disciplines, such as smart health services delivery, smart traffic management, border management, and governmental control. There is no single standard for IoT technology; thus, interoperability between IoT devices that use different protocols and standards is required. This research was carried out to provide and develop a specialized framework for an IoT-based smart health system by focusing particularly on interoperability problems. Based on different technology standards and communication protocols, the specific requirements of the IoT system were analyzed and served as a basis for the design of the framework. The protocols and standards within the framework utilize existing web technologies, communication protocols, and hardware design. This approach ensures that the specific expectations of the proposed model can be fulfilled with confidence. The experiments showed that interoperability between different IoT devices, standards, and protocols in a smart health system could be achieved using a specialized gateway device and that different web technologies could be used simultaneously in constrained and Internet environments.

1. Introduction

Many physical objects such as devices, vehicles, edifices, and other objects are used in traditional network infrastructure to define the Internet of things [1]. IoT technology uses the preestablished infrastructure of networks to ensure its validity. Many popular smart devices are utilized widely, including smart phones, tablets, and sensor-equipped devices [2]. Sensors in smart phones include the accelerometer, gyro, and proximity sensors. The acceleration of a body and the change in rotational angle are measured by using accelerometer sensors, the detection of nearly placed objects is measured with a proximity sensor, and body positioning is measured by GPS technology. IoT objects can be identified by using RFID tags [3]. The application list for uses of IoT technology is increasing day by day. There is also a research-based prediction that by the end of 2020 there will be 36.5 billion wireless connections, and 70% of wireless connections will consist of sensor devices while 30% will be without sensors [4].

Various IoT projects encourage the use of IoT technology, including body applications, the smart home, the smart city, and smart environment projects. In the smart home, protection and automation projects are involved. To create the smart environment, pollution, weather, earthquake and tsunami detection, and monitoring projects have been initiated. Smart city projects include smart transportation, country border security, electronic governance systems, smart city supply chain management, and grid station monitoring [5]. Due to the extensive use of IoT technology in every single field of life, its use in the health sciences is also natural. Therefore, different IoT-based smart health services projects are being initiated worldwide. Various types of smart health services are being provided to the public. These may include the following: remote monitoring of patient health, patient handling in an emergency, medication and routine health checkup reminders, remote patient prescriptions, and searches for the nearest health resources to the patient, such as doctors, paramedical staff, medicines, ambulance services, and many

other health resources. Health is an important entity for human life, and it has a great impact on the economy [6].

Many frameworks have been proposed to implement IoT technology in health services [7]. Three major paradigms have been discussed in the literature for IoT technology: application, security, and efficiency domains. The framework proposed focuses on the security, efficiency, and application domains of IoT technology in health services. This framework provides an overall design and implementation strategy with IoT technology. It addresses all the practical implementation issues of the technology (i.e., communication entities, communication technologies, hardware structure, data storage, data flow, and access mechanisms). A framework proposed by Zhao et al. [8] remotely monitored elderly people. They focused on the application domain of IoT technology. Their model facilitated the needs of elderly people. Machine learning techniques were utilized in their proposed model [9]. To monitor elderly people and patient status, another framework, Help to You (H2U), was proposed by Basanta et al. [10].

In the application paradigm, the importance of remote monitoring for patients was presented by Swiatek and Rucinski [11]. Their proposed model emphasized distributed system in delivering smart health services. They also addressed the innovative and commercial aspects of e-health services. Yang et al. [12] merged the traditional concept of a medical box with smart health services. iMedBox, iMedPack, and Bio-Patch are being used to provide medical services. An efficient resource-optimized rehabilitation system in IoT environments was proposed by Fan et al. [13]. Semantic information was utilized in their proposed model to efficiently identify the medical resources of a smart health system.

To keep the electronic health record and location information confidential, a specialized framework was proposed by Ding et al. [14]. They worked on the security paradigm of IoT technology. However, their main focus was limited to the security of personal location, personal identification, and identification of queries and personal electronic health records. Gong et al. [15] proposed and implemented a lightweight algorithm for the smart healthcare system. A lightweight private homomorphism was proposed, in addition to modified encryption DES algorithms. Improved algorithms provide the confidentiality needed for electronic health records during communication and while residing on the server. Various researchers have shown the use of IoT environments in the medical field [16–18].

2. Methodology

Different frameworks have been proposed to implement IoT technology in smart health systems. A specialized framework is required to address technical issues, such as interoperability and constrained and open Internet environments, as well as to address the nontechnical aspects, such as smart health services at the door step and remote consultancy for the poor people of underdeveloped countries. In this paper, we model a specialized architecture to provide smart health services in a smart health unit by using a specialized IoT gateway, which provides the interoperability between different sensor-based communication devices and provides the translation between local and Internet traffic. The IoT gateway also provides connectivity with backend cloud services.

The model presented also uses the constrained application protocol in the constrained environment and the hypertext transfer protocol in the Internet environment, due to the different requirements of both environments. The proposed framework also used the JSON format to store the list of remote consultants on the cloud, which was verified by the governmental health authorities. Only the physicians at local health centers who are registered practitioners can use the list verified by the health authorities. This approach ensures that only qualified practitioners at local health centers and remote consultants can use the list.

3. Research Scope and Validation Details

To summarize, the key goal of this paper is to present a specialized framework for an IoT-based smart health system. The framework uses a layered approach to address key issues of IoT-based smart health systems and provides a complete mechanism of data collection from the patient to cloud storage, which can be accessed either locally or remotely. The proposed model was tested using the Contiki real-time operating system, which utilizes the cooja simulation tool to simulate the behavior of network.

4. Background

In this section, we present background information for developing a better understanding of the proposed architecture.

4.1. IoT Devices. IoT devices are small in size, operate on a low power supply, and have limited processing capacity. Microcontrollers with 8-bit and 16-bit processors are well known in the market. A specialized foreground-background algorithm is used for a single processor to manage multiple processes at a time. Processors with 8-bit or 16-bit architecture are not specialized for planarity to support IoT devices, and these devices place demands on a real-time operating system. A real-time operating system requires more energy and memory and higher processing capabilities for working with devices. There are also many other issues for devices to work in the IoT paradigm. Devices must support TCP/IP stack for networking, but this stack is not a simple program, because more RAM is required to handle the number of network buffers for TCP. Furthermore, Java support is also a demand for an IoT device; therefore Java Virtual Machine (JVM) should also be run on the operating system of IoT devices. In conclusion, RAM and ROM support should be available for IoT devices to support the real-time operating system and communication stack.

Today, 32-bit microcontroller units that are small in size, operate on a low power supply, and have sufficient processing capacity to support IoT devices are also available on the market. These 32-bit microcontroller units are the best selection for IoT solution developers and providers.

Table 1: Wireless technologies for IoT systems.

Standards	Operating Frequency	Data Rate	Range	Power Consumption	Battery Time
IEEE 802.15.4	868/915 MHz, 2.4 Gz	250 kbps	10 to 300 m	Very Low	Months-year
Wi-Fi	2.4 to 5.8 GHz	11–105 Mbps	10 to 100 m	High	Hours
Bluetooth	2.4 GHz	723 Kbps	10 m	Very Low-Low	Days-Weeks

Table 2: Web technologies for IoT systems.

Protocol	Transport Mechanism	Messaging Method	Resource Consumption	Successful Applications
HTTP RESTful	TCP	Request/Response	10 Ks Flash or RAM	Smart home and grid
CoAP	UDP	Request/Response	10 Ks Flash or RAM	Used in Field Area Networks (FAN)
MQTT	TCP	Request/Response Public/Subscriber	10 Ks Flash or RAM	Remote monitoring and controlling of devices
XMPP	TCP	Request/Response Public/Subscriber	10 Ks Flash or RAM	Remote management of major appliances (white goods)

Data acquisition, processing, power management, communication stack, protocol conversion, firmware upgrade, and customizable security features can be implemented in IoT devices by using 32-bit microcontroller units. Intel and ARM families are well known in the market of 32-bit architecture processors.

The Intel family provides support for industrial Internet applications with atom processors, while the Intel Quark also captures the embedded system's market. On the other hand, the ARM Cortex-M0 processor is specialized to provide a low-cost product for IoT systems.

ARM Cortex-M3, M4, and M7 are the best choices to build IoT gateway devices. High performance, energy balance, and flexible system interfaces are major attributes of these processors for supporting IoT gateways. To support low energy consumption and high performance, the RL78 is the best 16-bit processor in a new generation of Renesas microcontrollers.

Different competitors for microcontroller units are providing their market solutions with pros and cons, but to support smart and small-embedded systems, ARM, Intel, and Renesas are popular and well tested. To support small IoT resource-constrained devices, Oracle's Java ME embedded 8 has also been designed. Java ME embedded 8 is widely used in wireless modules, buildings, industrial controls, health systems, grid monitoring, and many other applications.

Java ME embedded 8 requires that the system is based on the ARM architecture system-on-chips (SOCs), has only 128 KB of RAM and 1 MB of ROM, has a simple embedded kernel or operating system, and has a network connection that is either wired or wireless [19]. Java ME embedded 8 is also among the low-cost solutions and is sufficient to support resource specific devices of an IoT system.

4.2. Wireless Technologies for IoT. A standard communication technology is required for an IoT system. For communication between IoT devices and backend service providers, a communication technology should be chosen. IoT devices are enabled with wireless connectivity. There is no single standard wireless technology to support an IoT system. A number of technologies that have pros and cons are available. Implementation of wireless technology also depends upon the IoT project. For example, it may be a healthcare or home automation project, or a smart grid or environment-monitoring project. A comparison of some wireless technologies is given in Table 1.

4.3. Web Technologies for IoT. Existing web advancements can be used to develop IoT systems. However, these advancements are not sufficient to properly support IoT systems; therefore, results are poor. JSON and XML can be delivered in payloads by using HTTP and WebSocket protocols. Existing web protocols can be implemented on IoT devices, but these protocols require more resources to support IoT applications. To support IoT systems, many specialized protocols have been developed that can work efficiently with resource-constrained IoT devices and networks. Some web technologies are presented in Table 2.

4.4. System Components. This section presents the various components of the proposed system and further outlines the proposed model with the functionality of each layer.

The proposed model defines the structure of the IoT-based smart health system. The data collector, IoT gateway, backend facilitator, and access applications are the major components of the model. Fundamental parts of the proposed system are described below.

4.4.1. Data Collector (Dc). This component of the IoT system is used to sense the patient body. Data collectors are the sensor devices. These sensors monitor the health state and convert it to digital values. Data collectors support various types of wireless communication technologies to communicate with IoT devices.

4.4.2. IoT Gateway (iGW). The IoT gateway is a key component of the IoT system that connects the local processing units with the remote backend facilitator by using the Internet protocol IPv4 or IPv6. The IoT gateway is also used to convert the protocols, manage the IoT devices, and provide temporary storage. It is also a middle entity between the local sensor network and remote IP network. The IoT gateway also acts as middleware. It supports different modules to provide different functionalities in the IoT system.

4.4.3. Backend Facilitator (Bf). The backend service provider is the backend facilitator. These services may be outsourced from a third party or may be their own deployment. The backend facilitator provides storage services to permanently store the IoT data and perform decisions and analytics on the stored data. The data integration facility integrates the different types of data. Tools for security management and application development are also part of the backend facilitator. Additionally, remote consultancy is also linked with backend services, since the backend facilitator manages the list of remote consultants.

4.4.4. Access Applications (AA). The final requirement of the IoT system is the access mechanism for the IoT services, which is accomplished by using the access application. The access application may be installed on smart devices or on desktop systems.

4.5. Proposed Layers. Three layers that address the complete functionality of the system have been proposed for the IoT system. This part of the research discusses the functionality of each proposed layer. Each layer is utilized to provide smart health services. Different protocols and standards are used by the components of each layer to carry out their respective functionality. The proposed layers are also helpful for understanding the functionality of different components of the IoT system.

The three layers of the proposed model are described as follows:

(1) Sensor layer

(2) Network access layer

(3) Service access layer

The *sensor layer* is the first layer of the proposed model that addresses the functionality of the various components. Data collectors in this layer are used to monitor and accumulate the health information of a patient. The data collectors are the sensor devices, which are sometimes embedded in the body or may sometimes reside on the body of a patient. Data that are detected by the data collectors may include the pulse rate and heartbeat. A sensor device supports different communication technologies, as these may be from different vendors.

The sensor layer includes the following components:

(i) *Communication technologies*

(ii) *Bar Code and RFIDs used for tagging*

(iii) *Data collectors such as sensor devices*

(iv) *IoT gateway device (iGW)*

Local communication technology is used to transfer the sensed information to the IoT gateway, and then the IoT gateway transfers the information to the backend facilitator in IP format.

The *network access layer* is the second layer of the IoT-based health system and is used to provide connectivity between the backend facilitator and IoT gateway. This layer also provides an interface to the devices in the sensor layer with the backend facilitator (Bf). The cloud service provider supplies backend services. Dslam, DSL, and 3G/4G technologies are used to provide connectivity between the IoT gateway device and the backend services over the cloud by using Internet services. Specialized web technologies for IoT systems are also used to obtain the data.

The data collected in the smart health unit is forwarded to the backend cloud over the Internet by using the IoT gateway device. Many services are provided by the backend facilitator, such as data storage to store the data permanently, allowing querying by using query services, data integration of data from multiple sites by using data integration services, data analysis for future prediction, and many development tools provided by the backend service provider to develop new applications for the smart health system. An authentic list of remote consultants is also managed over the cloud storage. Government authorities authenticate consultants who are experienced, well known, and experts in their profession. Only registered practitioners in the smart health unit can use this list whenever they want to access their patient records.

The network access layer includes the following components:

(i) *Communication technologies*

(ii) *Backend cloud services*

(iii) *IoT gateway*

(iv) *A list for registered remote consultants.*

Thus, the network access layer is also an important part of the proposed model.

The *service access layer* provides access to the services of the smart health unit by using access applications. Applications may be installed on the smart computing devices or on desktop computers. The doctor or medical professional can access the health information of a particular patient whenever it is required. Local medical professionals in the smart health unit can consult the remote consultants by using the special application in the smart health unit.

Different application protocols are used to access the patient data from cloud storage. Data are also queried by using specialized database applications that use specific application protocols. IoT devices can also be accessed locally or remotely for management purposes. Remote consultants can treat the patient remotely from anywhere in the smart health unit of the IoT-based smart health unit.

The IoT system works in a constrained environment and in an Internet environment. The smart health unit part works in the constrained environment, where data collectors

and the IoT gateway are available. Therefore, there should also be state-of-the-art web technologies to work in this environment. Fortunately, CoAP is the best choice for working in a constrained environment. CoAP works best with constrained devices and constrained networks. However, there is no need to use CoAP in the Internet environment. HTTP RESTful is the best choice for working in the Internet environment. In the Internet environment, much network bandwidth is available that has high processing capacity for network devices. HTTP and CoAP can be mapped by using proxy services on the IoT gateway device.

The service access layer includes the following components:

(i) *Medical professionals*
(ii) *Management authorities*
(iii) *Smart computing devices and desktop computers*
(iv) *Smart applications*
(v) *Modern web technologies*

The service access layer in the smart health unit provides the front-end interface for its users.

4.6. Security Model. To secure the IoT system, first the devices should be secured from physical access by using locked racks. The wired connections and device features, which are not required, should be disabled. Default passwords should also be changed, and strong passwords should be used on devices. The remote access should be restricted when there is no need, and updates should only be installed when they are available. Additionally, the vendor's security measurements on the device should be assessed before purchase. Where possible, devices enabled with a self-correcting mechanism should be purchased. There must be a strong authentication mechanism with the cloud to protect against misidentification of the device. A strong encryption mechanism during communication can protect the data against illegal access. TLS must be used with the certificate validation mechanism for authentication and secure key distribution.

Communication between mobile applications and devices is carried out on a wireless connection; therefore, data should be encrypted during communication; otherwise local traffic will be unveiled. The mobile application should use TLS/SSL and validate the device's TLS certificates, which will protect the communication against a man-in-the-middle attack. Communication between mobile or web applications and cloud services should be secured with TLS/SSL by allowing its use from the cloud service; otherwise the attacker will capture the data passively. Many cloud service providers allow users to create a weak password that is not secure.

The service should enforce strong password creation. Strong passwords increase the effort needed to crack them when brute force or dictionary attacks are used. Mobile applications should follow best practices and be designed to work securely with services. Applications should properly validate the server's TLS certificate. Thus, best practices lead to secure communication. Figure 1 shows the security model for the IoT-based health system.

4.7. Modeling Exercise. Mathematically, the smart health unit can be presented as follows:

Generally, for the smart health unit,

$$\mathrm{SHU} = f\,(\mathrm{Dc}, \mathrm{iGW}, \mathrm{Bf}, \mathrm{AA}) \tag{1}$$

In (1), the smart health unit (SHU) variables are defined as follows: f represents function; Dc represents data collector devices; iGW represents the IoT gateway in the smart health unit; Bf represents the backend facilitator, which provides the backend services for the smart health system; and AA represents the access application. Equation (1) describes the complete smart health unit that depends on every component of the proposed model.

Equation (2) describes the smart health unit as follows:

$$\mathrm{SHU} = \left(\alpha + \beta\mathrm{Dc} + \gamma\mathrm{iGW} + \delta\mathrm{Bf} + \theta\mathrm{AA} + \mu^{\circ}\right), \tag{2}$$

where, for each component,

$$\mathrm{Dc} = f\left(\text{data collection}, \text{data forwarding}\right) \tag{3}$$

Dc is a function that represents the data collection and is used to accumulate the data and further forward it to the IoT gateway device for processing.

Now specifically,

$$\mathrm{CT} = f\,(\mathrm{iGW}), \tag{4}$$

where

$$\mathrm{CT} = (\text{Bluetooth, Wi-Fi, Zwave, 6LowWPAN, DSL, 3G, 4G}, \ldots, n) \tag{5}$$

The component CT represents the communication technologies that are supported by the IoT gateway device, iGW. Communication technologies may be Bluetooth, Wi-Fi, 6LowWPAN, DSL, or 3G/4G technologies.

4.8. Implementation Details. To access the health data locally or from cloud storage, state-of-the-art smart access applications are used. HTTP is an application layer protocol that is well known for web technologies, and it is used to retrieve and store the data over the backend cloud storage within the Internet environment. HTTP works with the transmission control protocol TCP, which is a transport layer protocol. To access the web resources, HTTP uses defined methods, such as "GET" to get the resource, "PUT" to put the web resource, and many other methods, such as "POST" and "DELETE".

When a particular client establishes a connection with the server, HTTP, which is a connection-oriented protocol, uses these defined methods after establishing the connection. The TCP 3-way handshake is a connection-oriented mechanism that is used to establish the connection from a client to the server. The HTTP request is sent from the client to the server, and the connection is established between the client and the server after the handshake process.

For example, if the body temperature of a patient is a resource, HTTP will use its "GET" method to access this resource. HTTP uses a universal resource identifier (URI) to

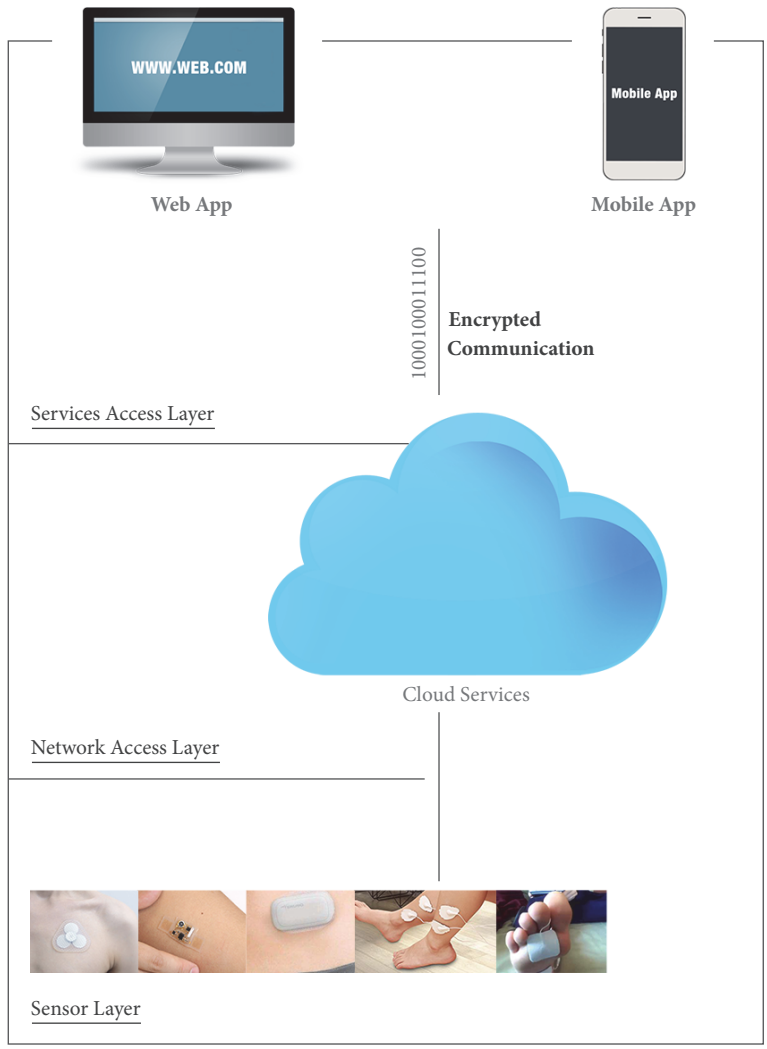


Figure 1: Security model.

identify the specified resource. Since HTTP is connection-oriented, it will also require the connection termination mechanism to terminate the connection. The termination process is also carried out by a 2-way TCP termination process. Once a connection is terminated, a connection establishment process will be needed to access the resource again. Establishing and terminating the HTTP connection process is shown in Figure 2.

A simple representational state transfer in the RESTful architecture is used by the HTTP. Predefined sets of operations are provided for its simple work. The XML, HTML, or JSON format are used to represent the resources in response to the RESTful request.

HTTP and CoAP are both used in the smart health system. In the constrained environment, only CoAP is used since it is specialized to work in constrained networks over constrained devices, and in the Internet environment, HTTP is used. Mapping between the HTTP and CoAP environments is performed on the IoT gateway device by using a special proxy module. The mapping is shown in Figure 3.

In order to access the web resource, HTTP is used in the Internet environment, which is based on a variety of networks. Networks have high bandwidth; therefore there is no issue of resource consumption as in constrained environments. Only CoAP is implemented in the constrained environment, as it is specially designed to work with constrained devices.

The response code is used to determine the proxy/caching model that is supported by the CoAP [20]. The CoAP-HTTP proxy model provides efficiency in the constrained environment where IoT devices work. If a resource named "heartbeat" of a patient is accessed using the HTTP request, then this request is generated in the Internet environment. The IoT gateway device responds to the HTTP request that is supporting the proxy model. The IoT gateway converts the HTTP request to the CoAP request and then sends the response back to the HTTP request. Figure 4 shows how the CoAP-HTTP proxy model works.

The CoAP, as a built-in service, also supports the subscription mechanism. The CoAP supports a special option

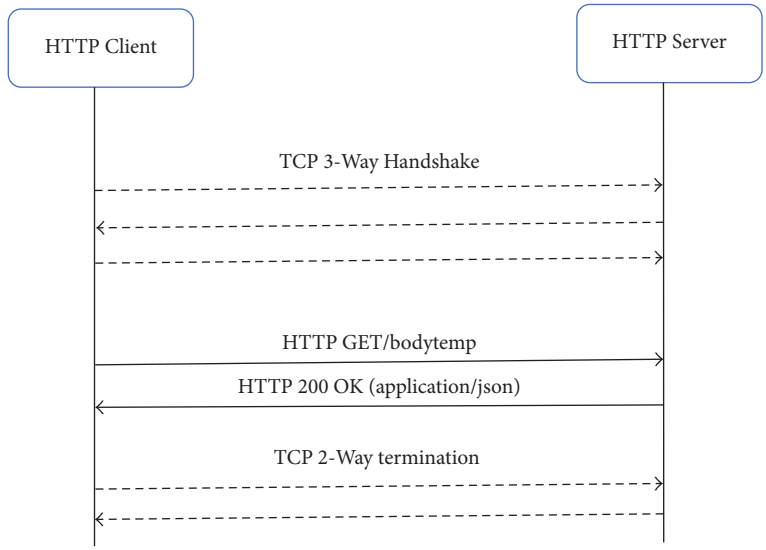


Figure 2: HTTP connection establishment and termination.

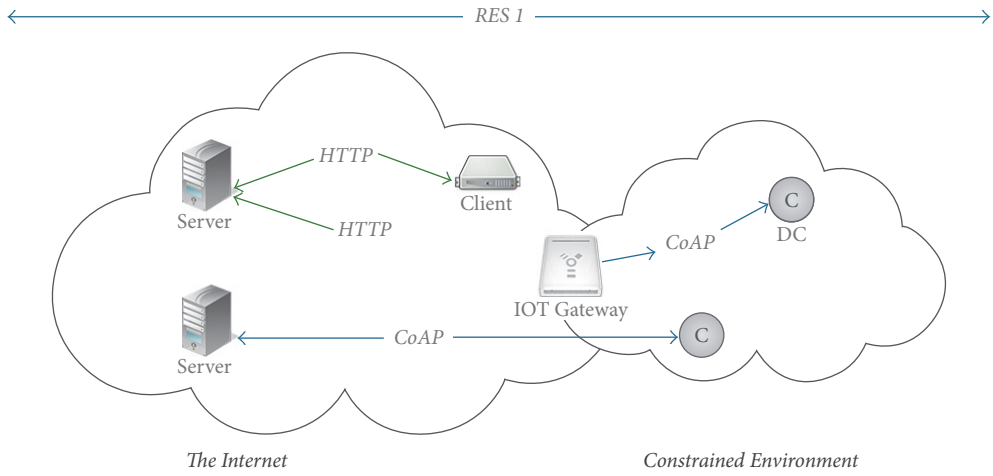


Figure 3: Mapping between CoAP and HTTP.

named "observe" to support the subscription method. Figure 5 shows the subscription-based "GET" method.

Different observations are separated by using tokens and therefore remain identified. When the value of the observation is changed, the server responds with a new observation but with the same token. The client also acknowledges when it receives the observation.

A lightweight data interchange format known as JSON is used to store the remote consultants list. It is a good format for the timely delivery of the message between the access entities and the service providers, for example, the backend facilitator Bf and the access application AA. The remote consultants list is accessed using the HTTP protocol, since the list is part of a web resource on the cloud storage [21]. A list in JSON format is shown in Algorithm 1.

The list of remote consultants that is stored on the cloud storage can be accessed by the authorized consultants in the smart health unit whenever they require it. A list is shown in Algorithm 1 that shows their name, age, gender, location, contact information, and time availability, so that the physician in the smart health unit can determine the appropriate remote physician. The list shown in Algorithm 1 provides the details of a physician, a pediatrician, and a cardiologist. The remote consultants list is also verified by the higher government authorities, so that only registered practitioners can participate in the IoT-based smart health system. This is compulsory; otherwise nonexperienced or even nonprofessional persons can register themselves inappropriately for the wrong purposes.

4.9. Experimental Results. An experiment was conducted using the Contiki operating system with 8 sensor nodes in a medical unit. The Contiki operating system also has a cooja simulating tool that monitors the behavior of network. The following results show the behavior of a network at the proposed sensor layer. The results show the behavior of the

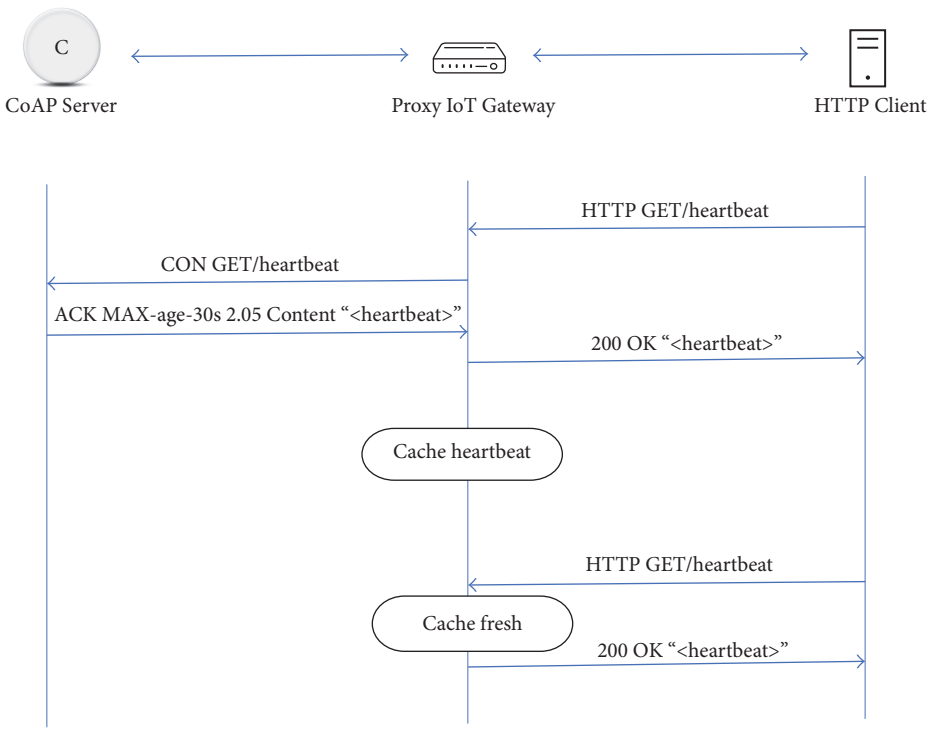


Figure 4: CoAP-HTTP proxy using the GET method.

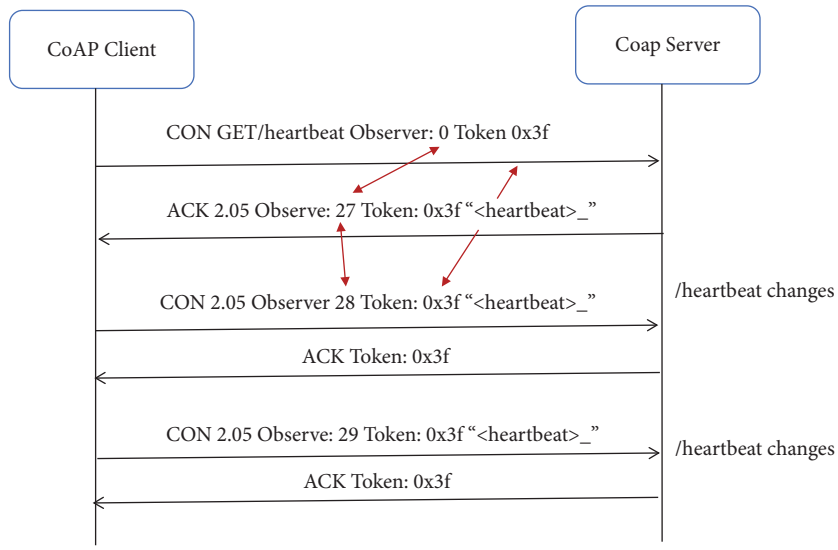


Figure 5: CoAP subscription-based GET method.

constrained environment in which smart IoT devices and gateways work, where collecting, processing, and forwarding the sensed data occur. Since IoT devices work in a constrained environment, there is a limit to the data rate and usage of sensor devices. Wireless terminals have a limited capacity to send and receive data, similar to this experiment. The maximum data rate, as given by the IEEE 802.15.4 standard, is 250 kbps [22]. Figure 6 shows the bandwidth consumption of the data collectors in the smart health unit.

The results show that the data rate varies with respect to the work of all of the nodes. However, it is not necessary that all devices must work at the same time. There is the possibility that, at a given time, some devices may be sending and receiving data, but other devices may not be doing so. Only the number of devices and the bandwidth do not measure the data rate. Protocols, which are being deployed for communication, and the network topology are also concerning factors for measurement of the efficiency of a system. The IoT gateway device's capacity is also a factor in

```
var consutants = {
    "physician" : {
      "name" : "Ijaz Malik",
      "age" : "50",
      "gender" : "male"
      "location" : "NH Multan"
      "contact" : "03xxxxxxxxx"
      "availability" : "Morning"
      "mail_id" : "exmple@gmail.com"
},
 "pediatrician" : {
      "name" : "Fawad Bukhari",
      "age" : "45",
      "gender" : "male"
      "location" : "BVH BWP"
       "contact" : "03xxxxxxxxx"
       "availability" : "Morning"
       "mail_id" : "exmple@gmail.com"
},
"cardiologist" : {
      "name" : "Aftab Ahmmad",
      "age" : "38",
      "gender" : "male"
      "location" : "CPEIC Multan"
       "contact" : "03xxxxxxxxx"
       "availability" : "Morning"
       "mail_id" : "exmple@gmail.com"
  }
}
```

Algorithm 1: Remote consultant list stored in JSON format.

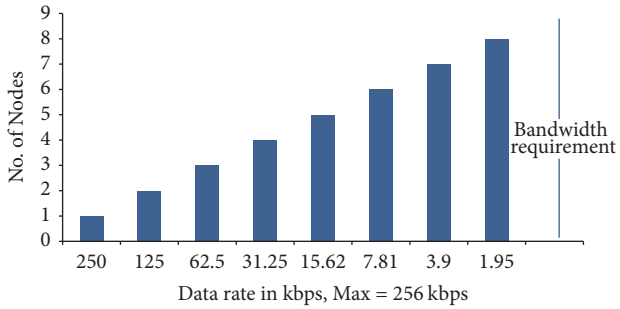


Figure 6: Bandwidth consumption by the data collectors.

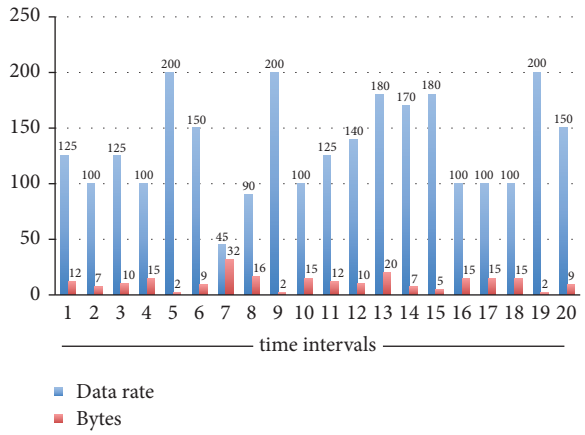


Figure 7: Data rate at different intervals of time.

understanding the proper working of the sensor devices since the IoT gateway directly interacts with the sensor nodes.

At different time intervals, the data rate is different for different devices in the smart health unit. Whenever there is a high amount of data to transfer, then the maximum data rate will be low, and whenever there is a low amount of data to transmit or receive, the data rate will be high. Figure 7 clearly shows the diversion in the data rate.

4.10. Implementation Scenario. A concept of the proposed model is illustrated in a particular smart healthcare scenario in Pakistan.

For example, people in Pakistan live mostly in villages, and when a patient in a village requires his medical checkup, he visits the nearest smart health unit in his village. In the smart health unit, all of the necessary equipment is installed to provide smart health services.

The smart health unit is equipped with data collector (Dc) sensor devices to sense the health information of the patient, such as body temperature, heartbeat, respiratory rate, and glucose level, and an IoT gateway to act as a middle entity between the local sensor network and the backend facilitator. The smart health unit is also equipped with the smart computing device and desktop computers, which will be used by the medical staff in the smart health unit to access the data and IoT devices. There is also an Internet connection to provide connectivity between the smart health unit and the backend facilitator, using either the mobile network GSM, 3G/4G, PSTN, ADSL, or a DSL connection.

When the patient asks for his medical checkup, the staff offers him wearable sensor devices that connect the patient to the devices of the medical staff. Wireless sensor devices may use Bluetooth, 6LowWPAN, ZigBee, Z-Wave, and many other wireless technologies. Data collectors process the sensed information to the IoT gateway device, which has temporary storage for the sensed data, and these data become accessible to the medical staff in the unit by using smart computing devices or desktop computers. The IoT gateway is smart enough to support multiple wireless technologies, which solves the interoperability issue between the sensor devices and the gateway device in the smart health unit. The IoT gateway device is also capable of formatting the sensed data to the Internet format, which supports the use of the Internet for communicating with the backend facilitator.

The smart health unit data are forwarded to the IoT gateway device and to the backend facilitator, that is, the cloud service provider. An authorized person from anywhere in the world can, at any time, access the data over the cloud. Since there is also an authorized consultant list in the cloud storage, the medical staff in the smart health unit can use it to consult with the remote consultants. Data using the cloud storage is permanently stored, so the patient data can be accessed whenever it is required. These data are also

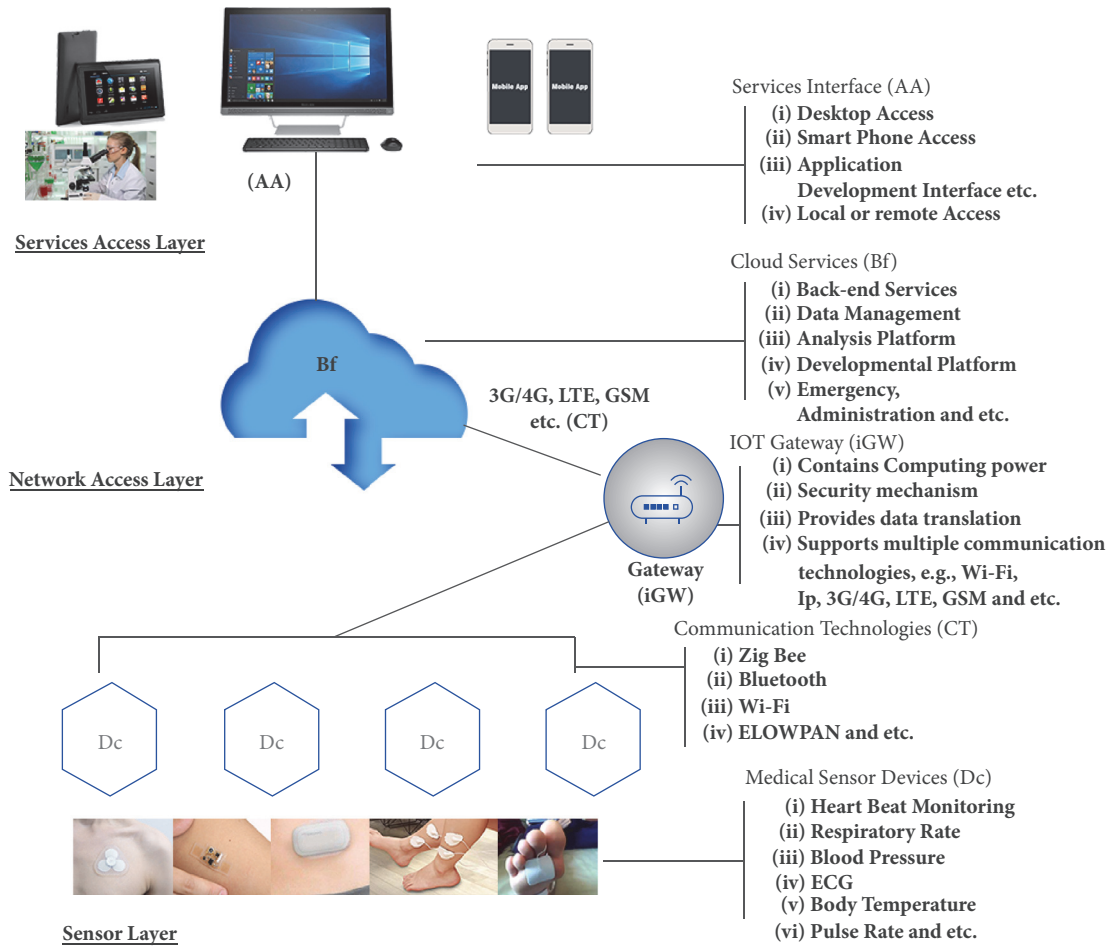


Figure 8: Smart health scenario.

accessible to government authorities to keep track of public health. Therefore, the patient in the smart health unit is taking advantage of health services at the doorstep. The medical staff is also taking advantage of the remote consultants to provide better health services. An IoT-based smart health unit scenario is shown in Figure 8.

5. Conclusions

Over the past few years, the scope of IoT technology has widened in every field of life. Different research problems have been raised in the deployment of IoT technology. To mitigate the problems and find better deployment solutions, different work has been performed in the research community. In the field of health sciences, different frameworks have been proposed to efficiently deploy the smart health services, and the focus has always been on the security and efficiency of these algorithms.

A specialized framework is presented in the current research that provides smart health services in underdeveloped countries, especially in rural areas. The framework studies various aspects of IoT technology for smart health services, such as the interoperability and standardization issues, constrained and Internet environments, specialized communication protocols, and web technology requirements. The proposed model consists of three layers, where each layer performs a specialized task. In the future we aim to develop a detailed security infrastructure that can be incorporated using the current framework.

Conflicts of Interest

The authors declare that there are no conflicts of interest related to this paper.

References

[1] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton, "Smart objects as building blocks for the internet of things," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2010.

[2] C. Zhu, V. C. M. Leung, L. Shu, and E. C. Ngai, "Green Internet of Things for smart world," *IEEE Access*, vol. 3, pp. 2151–2162, 2015.

[3] E. Mok, G. Retscher, and C. Wen, "Initial test on the use of GPS and sensor data of modern smartphones for vehicle tracking in dense high rise environments," in *Proceedings of the Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS '12)*, pp. 1–7, IEEE, October 2012.

[4] "Internet of things vs. Internet of everything what is the difference, May 2014".

[5] J. Haase, M. Alahmad, H. Nishi, J. Ploennigs, and K. F. Tsang, "The IOT mediated built environment: A brief survey," in *Proceedings of the 14th IEEE International Conference on Industrial Informatics, INDIN 2016*, pp. 1065–1068, IEEE, July 2016.

[6] M. García, "The impact of IoT on economic growth: A multifactor productivity approach," in *Proceedings of the International Conference on Computational Science and Computational Intelligence, CSCI 2015*, pp. 855-856, IEEE, December 2015.

[7] M. A. Sahi, H. Abbas, K. Saleem et al., "A Survey on Privacy Preservation in e-Healthcare Environment," *IEEE Access*, 2017.

[8] W. Zhao, C. Wang, and Y. Nakahira, "Medical application on internet of things," in *Proceedings of IET International Conference on Communication Technology and Application (ICCTA 2011)*, pp. 660–665, Beijing, China, 2011.

[9] S. Earley, "Analytics, machine learning, and the internet of things," *IT Professional*, vol. 17, no. 1, Article ID 7030173, pp. 10–13, 2015.

[10] H. Basanta, Y. P. Huang, and T. T. Lee, "Intuitive IoT-based H2U healthcare system for elderly people," in *Proceedings of the Networking, Sensing, and Control (ICNSC), 2016 IEEE 13th International Conference on*, pp. 1–6, IEEE, 2016.

[11] P. Swiatek and A. Rucinski, "IoT as a service system for eHealth," in *Proceedings of the e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pp. 81–84, IEEE, Lisbon, Portugal, October 2013.

[12] G. Yang, L. Xie, M. Mäntysalo et al., "A health-IoT platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2180–2191, 2014.

[13] Y. J. Fan, Y. H. Yin, L. D. Xu, Y. Zeng, and F. Wu, "IoT-based smart rehabilitation system," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1568–1577, 2014.

[14] D. Ding, M. Conti, and A. Solanas, "A smart health application and its related privacy issues," in *Proceedings of the Smart City Security and Privacy Workshop (SCSP-W)*, pp. 1–5, 2016.

[15] T. Gong, H. Huang, P. Li, K. Zhang, and H. Jiang, "A medical healthcare system for privacy protection based on IoT," in *Proceedings of the Parallel Architectures, Algorithms and Programming (PAAP), 2015 Seventh International Symposium on*, pp. 217–222, December 2015.

[16] K. R. Amrutha, S. M. Haritha, V. M. Haritha, A. J. Jensy, S. Sasidharan, and J. K. Charly, "IOT based Medical Home," *International Journal of Computer Applications*, vol. 165, no. 11, 2017.

[17] M. M. Rathore, A. Ahmad, A. Paul, J. Wan, and D. Zhang, "Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health," *Journal of Medical Systems*, vol. 40, no. 12, article no. 283, 2016.

[18] G. Zhang, C. Li, Y. Zhang, C. Xing, and J. Yang, "SemanMedical, A kind of semantic medical monitoring system model based on the IoT sensors," in *Proceedings of the e-Health Networking, Applications and Services (Healthcom), 2012 IEEE 14th International Conference on*, pp. 238–243, IEEE, 2012.

[19] Java Embedded Documentation, http://www.oracle.com/technetwork/java/embedded/javame/embedme/documentation/index.html.

[20] Z. Shelby, K. Hartke, and C. Bormann, "The constrained application protocol (CoAP)," Internet Engineering Task Force, RFC 7252, 2014.

[21] D. Crockford, "The application/json Media Type for JavaScript Object Notation (JSON)," RFC Editor RFC4627, 2006.

[22] "IEEE 802.15 WPANTM Task Group 4," http://www.ieee802.org/15/pub/TG4.html.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6709607, 11 pages
https://doi.org/10.1155/2018/6709607

Research Article

Group Recommendation Systems Based on External Social-Trust Networks

Guang Fang, Lei Su, Di Jiang, and Liping Wu

Kunming University of Science and Technology, Kunming 650051, China

Correspondence should be addressed to Lei Su; s28341@hotmail.com

Received 14 February 2018; Accepted 16 April 2018; Published 29 May 2018

Academic Editor: Huimin Lu

With the development of social networks and online mobile communities, group recommendation systems support users' interaction with similar interests or purposes with others. We often provide some advices to the close friends, such as listening to favorite music and sharing favorite dishes. However, users' personalities have been ignored by the traditional group recommendation systems while the majority is satisfied. In this paper, a method of group recommendation based on external social-trust networks is proposed, which builds a group profile by analyzing not only users' preferences, but also the social relationships between members inside and outside of the group. We employ the users' degree of disagreement to adjust group preference rating by external information of social-trust network. Moreover, having a discussion about different social network utilization ratio, we proposed a method to work for smaller group size. The experimental results show that the proposed method has consistently higher precision and leads to satisfactory recommendations for groups.

1. Introduction

In recent years, the research of recommendation system and mobile computing has developed very rapidly, and many kinds of recommendation systems have appeared, for example, mobile recommendation system, context-aware recommendation system, and social network recommendation system. However, most of the current system can only be worked for a single user. In fact, many daily activities are carried out by a crowd of people from different regions, such as watching movies or TV programs, going to restaurant for meal, and traveling and getting service in public. Therefore, it is necessary for the system to consider suggestion of a certain number of people, which is called GRS (group recommendation systems) [1].

In GRS, group members' preferences may be similar or different. How to get the common preference of group members, alleviate the conflict among group members and make the recommendation result as much as possible to meet the needs of all group members, which is the key problem to be addressed [2, 3].

The current social network group recommendation systems consider both the strength of the relationship between the members of the group into account [4–6] and the influence of social network information on each group members [7–9] and finally generate group recommenders through aggregation strategies. At present, the main influence of social network is the social impact of the group members on the group recommendation systems. At first, when the group preferences are inconsistent, those systems take care of preference of the members with more social influence and ignore some intentions of the members with small social influence. In addition, it is unreasonable that the present GRS still consider the influence of social networks while the group has reached consensus. This paper aims at correcting preference rating by social-trust networks when group rating of item cannot reach consensus. Specifically, this paper uses the real ratings of the external members trusted by the group members to correct the predicted rating of some item. When the disagreement of group is small, namely, group tendency reaches consensus, the influence would be reduced by social-trust networks on GRS, so as to dynamically adjust the influencing factors of social networks and reduce the error on GRS.

In addition, whole group members are asked to be in the same social network on present social network

recommendation system [4, 5, 7–9], and the members of other social networks could be ignored. The method proposed in this paper is that some users who dislike communicating with others on social network and are not in the same social network can benefit from this as well.

The structure of this paper is as follows. In Section 2, we introduce the related work recommended by the social network, and Section 3 elaborates the method based on external trust social network proposed in this paper. Section 4 introduces the experimental results. Section 5 summarizes the full text and discusses future work.

2. Related Work

2.1. Group Recommendation Systems. GRS [3] usually generate group preferences through aggregating individual's ratings. According to the paper by Jameson and Smyth [1], the main approaches to generate the preference aggregation are merging the recommendations made for individuals, aggregation of ratings for individuals, and constructing a group preference model. Supan et al. [10] present a compilation of the most important preference aggregation techniques. These basic approaches merge the ratings predicted individually for each item to calculate a global prediction for the group. The selection of a proper aggregation strategy is a key element in the success of recommendation. The work by Ricci et al. [11] describes a series of experiments that were conducted with real users in order to determine which strategy performs best. These experiments show that the average and the average without misery strategies perform best from the users' point of view because they seem to obtain similar recommendations to those that emerge from an actual discussion in a group of "humans".

Group recommendation systems could be classified into two main categories [12, 13]: those which perform an aggregation of individuals' preferences to obtain a possible group evaluation for each candidate item and those which perform an aggregation of individuals' models into a single group model and generate suggestions based on this model. In the first method, an individual-based recommendation system is first used to generate recommendations for each group member, then a group consensus function is used to merge the individual recommendations and select ones that are most suitable for the whole group. In the second method, a pseudo user profile is generated from all group members, and an individual-based recommendation system is then used at run time to generate recommendations for the pseudo user. By considering the recommendations for individual group members and merging [7] them at run time to generate group recommenders, this GRS architecture can easily accommodate dynamic groups and tailor its recommendations for each specific scenario.

The individual recommender implements the Collaborative Filtering algorithm described in Kelleher and Bridge [14] and group recommendation has largely been studied in the context of Collaborative Filtering (CF) [8, 15, 16]. We have chosen this algorithm because it is widely used to recommend items when the modeling of user preferences is not a valid option (as in most of real scenarios [17–19] and others [20–23]). This algorithm requires users to rate an initial set of items. Then, those ratings are used to estimate the predicted rating for an unrated item.

Some instances have been applied on GRS, such as MusicFX [24] which obtains the information about the clients' interests from a database that stores their music genre preferences (previously and explicitly specified). Some systems make the suggestions which would not be particularly abrupt and maintain the consistency of the recommended style. In order to avoid abrupt music recommendation, music to be played would be considered in FlyTrap [25]. Another system that takes previous selections into account is PoolCasting [26]. It uses a Case-Based Reasoning system to generate a sequence of songs customized for a mobile community of listeners. To select each song in the sequence, first a subset of songs musically associated with the last song of the sequence is retrieved from a music pool; then the preferences of the audience expressed as cases are reused to customize the selection for the group of listeners; finally, listeners can revise their satisfaction (or lack of) for the songs they have heard. Let's Browse [27] first creates the individual profiles as a set of about 50 keyword-weighted pairs obtained by a crawler from the user's page. Then the group preference model results from a simple linear combination of the individuals' profiles.

In addition to personal information, there are some other pieces of information which would be taken into account [28, 29], and in another interesting content-based recommendation system there is Pocket Restaurant Finder [30], which recommends restaurants for groups of people based on user location and the culinary characteristics of the restaurant. McCarthy proposes the Pocket Restaurant Finder that recommends restaurants to groups of people considering their culinary preferences and location. Specifically, the recommender uses information like travel distance, expected facilities, cuisine desired, and budget planned. When using the recommendation system, the group members have to express explicitly and the desired values for the four features individually, and they also need to order the features in a level of priority. Then, Pocket Restaurant Finder computes the recommendation by applying an average preferences aggregation method. The restaurants are displayed in a ranked list that matches the group's likes.

2.2. Social Network Recommendation System. One of the important influencing factors of the GRS is the social relationship among group [8, 9, 31, 32]. Research shows that users prefer to accept the recommendation of trusted users rather than anonymous user [5].

Gartrell et al. [7] proposed that the social network merges into the group recommendation systems for the first time. The method is called Rule-Based Group Consensus Framework, which not only considers the group members' interest, but also describes the group members' weight difference. The system takes social relations, social frequency, and professional level into account.

With the development of the Internet, social platform has been focused by researchers as well. HappyMovie [8] is an online GRS on Facebook; moreover researchers need

to handle two tests: one is personality test, getting users' personality through TKI [33], and personality is divided from selfishness to tolerance, a total of five levels. Usually selfish users are not affected by others and tolerant users are easily impacted, in order to reach consensus in group, making vulnerable users change their recommendations. The other one is building user preference model through choosing enjoyed movies by themselves. Then group suggestion is generated through aggregating individual preferences.

SocialGR [9] is also a real research system, the system takes many aspects of social factors into account, and the main consideration is Trust Relationship (TR), Social Similarity (SS), and Social Centrality (SC). TR reflects the cohesion between two members by analyzing their affective relation. SS reflects the likeness between members, that is, shared activities, likes, friends, or interests. SC reflects members' reputations in the social network. The basic recommendation system utilizes a Collaborative Filtering recommendation system. And group recommender is determined through the Maximizing Average Satisfaction (MAS) strategy.

Current social networks and group recommenders usually considered the strength of the relationship among the members of the group; the tie among members would affect the results of GRS. Though some papers [4–7, 9] considered the social factors such as personality, trust, personal expertise, and social status of the members; once the information was acquired, the steps of the social networking group recommendation systems were as follows: At first, fetching the list of preference of all group members by individual recommendation system; then, increasing or decreasing the weight of a group of members of the article via social factors; finally, getting the ultimate results of group recommendation systems by some aggregation strategies, which consider the individual recommendation and weight. These methods utilize the social factors more or less; group members' weight is considered in ones; it means that a few of members would be ignored because of some group members play a decisive role, leading to the tendency to influence influential members. Therefore, this paper focuses on the impact of external members on the GRS, the information of external members, and internal members is taken as a whole, respectively.

3. External-Based Social-Trust Networks Group Recommendation Systems

3.1. GRS Calculation Framework. As to previous group recommendations in social networks, they consider the strength of the relationship between group members [4, 5] and the impact of social network information on each group member [7–9], hence generating group recommendation by the aggregation strategy. At present, the main impact of the social network on the group recommendation systems is the social influence of the group members. On the one hand, when the group preferences are inconsistent, those systems take care of preference of the members with more social influence and take no notice of some intentions of the members with small social influence; on the other hand, it is unreasonable that those system still consider the influence of social networks while the group has reached consensus. The idea of this paper includes the following: correcting preference rating through social-trust networks when group rating of item cannot reach consensus. Specifically, this paper uses the real ratings of external members that are trusted by the group members to correct the predicted rating of items. When the disagreement of group is small, namely, group tendency reaches consensus, the influence would be reduced on GRS by social network, so as to dynamically adjust the influence of social networks and reduce the error on GRS. In fact, most people hope to persuade others to follow their advices when faced with the conflict and large disagreement among the group. At that moment, if there is a role, which you trusted and could tell you what is the best choice, perhaps you would change your mind. Watching a movie with your friends, for instance, you may not intend to do it with them but someone you trusted would change your mind in that just he watched it before and has felt well. This scenario shows correcting group preference by new information when degree of group disagreement is large in GRS, and the group is able to reach consensus.

At first, fetching the list of preference of each group member by Collaborative Filtering, and ensuring a special group (by randomly selected group or fasten group), and getting the results of group recommendation systems by some aggregation strategies, the above progress is called classical group prediction. Then, calculating disagreement of item i by predictions of group members, disagreement is greatly affected by external trust networks and vice versa. The frame and formula of our method are Figure 1 and formula (1).

$$\text{gpred}(G,i) = \begin{cases} (1-\lambda_i)\cdot \text{CGP}(G,i) + \lambda_i \cdot \text{OAR}(\text{OG},i) & \\ \text{CGP}(G,i), & \text{if OAR}(\text{OG},i) = \varnothing \\ \text{OAR}(\text{OG},i), & \text{if CGP}(G,i) = \varnothing \end{cases} \tag{1}$$

Here, G refer to a group and OG (out of group) represents information of external group, namely, users that group members trusted. CGP(G, i) (classical group prediction) is predictions of item i by individual recommendation system and some aggregation strategies. OAR(OG, i) (outer actual rating) represents actual ratings of external members of group and λ_i is a dynamic balance factor of external members and internal members of item i, balancing ratings of

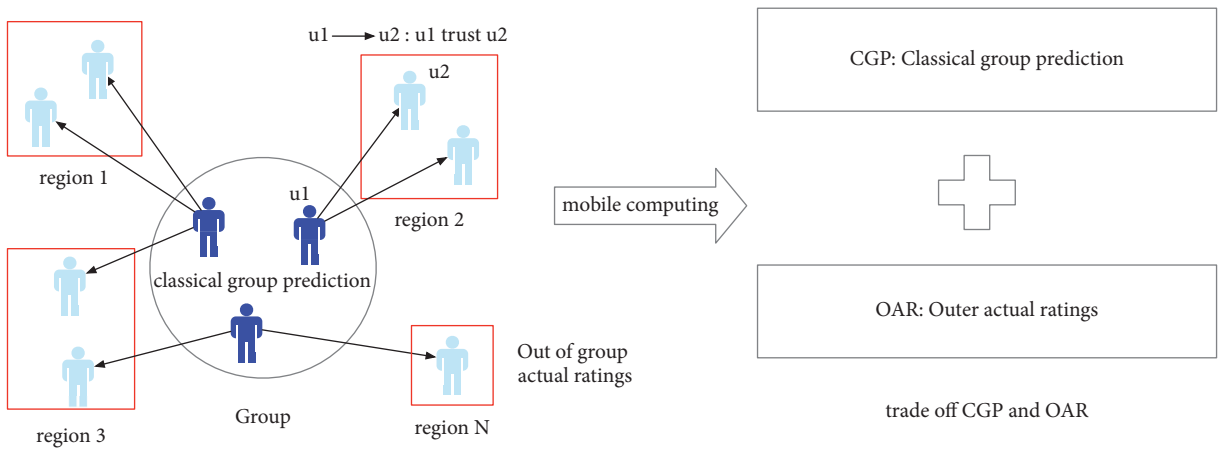


Figure 1: Frame of group recommendation systems.

recommendation between external and internal members. $\coprod$ represents a aggregation strategy. Note that, if $\mathrm{OAR}(\mathrm{OG}, i) = \varnothing$, then there are no real ratings of item i by trusted users, and results calculated by $\mathrm{CGP}(G, i)$ would be selected. Else if $\mathrm{CGP}(G, i) = \varnothing$, then there are no predicted ratings of item i by group users, and results calculated by $\mathrm{OAR}(\mathrm{OG}, i)$ would be utilized.

We have adjusted the intensity of the external influence dynamically. When the disagreement of group members is large, we need the external real score to correct. The bigger the disagreement is, the greater the external influence of group members' external influence would be. When the disagreement of group members is small, the opinions of the group members need to be retained. The smaller the disagreement is, the less the group members would be influenced externally. For example, when two people decide to watch a movie together, even if someone recommends a good movie, it may not be easy to influence the decision of the two. We need to solve two problems: one is how to determine the value of λ, and the other is how to solve $\mathrm{OAR}(\mathrm{OG}, i)$. $\mathrm{CGP}(G, i)$ (classical group prediction) represents predictions of item i by individual recommendation system and some aggregation strategies; the formula is as follows:

$$\mathrm{CGP}(G, i) = \coprod_{\forall u \in G} \mathrm{pred}(u, i). \tag{2}$$

$\coprod$ represents a aggregation strategy and $\mathrm{pred}(u, i)$ represents user u's predicted rating on item i by individual recommendation system.

In $\mathrm{OAR}(\mathrm{OG}, i)$, the actual ratings of external members, which are relative to group members, would be introduced into GRS by social network. Moreover, the object of each user trust may be more than one, and two points need to be considered on external actual information: degree of trust and obtaining actual rating set of members that are trusted; the formula is as follows:

$$\mathrm{OAR}(\mathrm{OG}, i) = \coprod_{\forall u \in G, v \in \mathrm{OG}} t_{uv} \cdot \mathrm{actual\_rating}(v, i). \tag{3}$$

Here, G refers to a group, OG (out of group) represents information of external group, namely, users that group members trusted, $\coprod$ represents a aggregation strategy, t_{uv} represents degree of trust of user u on user v, $t_{uv} \in [0, 1]$, 0 is trust scarcely, and 1 refers to trust completely. The more the influence, the high the degree of trust in external actual information; $\mathrm{actual\_rating}(v, i)$ represents actual rating of user v on item i.

3.2. Dynamic Adjustment of Parameter λ

3.2.1. LVD. This approach utilizes disagreement and refines a little on disagreement. Firstly, calculate the unbiased estimate of predicted ratings of group members, namely, sample variance in order to compute the number of points far away from the center in group preference set. The method is called LVD (lambda via disagreement), and the formula is as follows:

$$\mathrm{dis}(G, i) = \frac{1}{|G| - 1} \sum_{u \in G} \left[\mathrm{pred}(u, i) - \frac{\sum_{v \in G} \mathrm{pred}(v, i)}{|G|} \right]^2 \tag{4}$$

and the calculating formula of λ is as follows:

$$\lambda = \frac{1}{|G|} \sum_{u \in G} \mathbb{1} \left[\left(\mathrm{pred}(u, i) - \frac{\sum_{v \in G} \mathrm{pred}(v, i)}{|G|} \right)^2 > \mathrm{dis}(G, i) \right]. \tag{5}$$

Here, $|G|$ refers to the number of a group and indicator function is introduced: $\mathbb{1}[x]$ refers to that if x is true, then expression is 1, unless 0. $\mathrm{pred}(u, i)$ represents user u's predicted rating on item i by individual recommendation system.

3.2.2. LVTP. Owing to the structural of LVD is simple, some requirements should be met on the size and predicted distribution of groups. On the one hand, in the case with which

Table 1: λ via disagreement.

	User1	User2	dis	λ
Item_A	0	4	8	0
Item_B	0	5	12.5	0

Table 2: λ via disagreement.

	User1	User2	User3	dis	λ
Item_A	0	1	4	4.33	0.33
Item_B	0	1	5	7.0	0.33

only two members cannot cope, Table 1, the disagreements (dis) are 8 and 12.5, respectively, and the values of lambda are both 0, which means that adjustment is unnecessary on recommendations of external social network, and it is unreasonable. On the other hand, Table 2, the disagreements (dis) are 4.33 and 7.0, respectively, and the values of lambda are both 0.33; something looks well but still cannot reflect well with big disagreement on social network. Both of the above two cases with the large disagreement are observed, however having small values of lambda; as a result, this method cannot deal with the situation of mere few people.

In order to address the above issue, the method LVTP (lambda via two parts) is proposed in that case as the following steps: firstly, the standard value α_{ri}, a balancing item i, would be set; secondly, utilizing standard value α_{ri} divides the predictions of group members into different two parts, a big one is Greater_{ri} and a small one is Les_{ri}; finally, calculate lambda value on each item i. The formula is as follows:

$$\lambda_i = \frac{\mathbb{1}\left[\text{Greater}_{ri} \wedge \text{Les}_{ri} > \varnothing\right]\left(\overline{\text{Greater}_{ri}} - \overline{\text{Les}_{ri}}\right)}{\gamma + \max\left(\text{Greater}_{ri}\right)}. \quad (6)$$

Here $\overline{\text{Greater}_{ri}}$ represents the average value of bigger than standard value in rating set. Indicator function is introduced: $\mathbb{1}[x]$ refers to that if x is true, then expression is 1, unless 0. Note that, in step 3, if $\text{Greater}_{ri} = \varnothing$ or $\text{Les}_{ri} = \varnothing$, which means that ratings of group on item i reach consensus and indicates that group members are either like or dislike item i, then $\lambda_i = 0$, which means that social network has almost no influence on item i. Particularly, γ is a smooth factor, in order to avoid if $(\overline{\text{Greater}_{ri}} - \overline{\text{Les}_{ri}}) = \max(\text{Greater}_{ri})$ result in the GRS neglect of group's suggestions completely, as well as unreasonable.

With regard to the choice of standard values, this paper considers the following three aspects: (1) The mid value of the range can be evaluated; for example, the maximum score for a certain item is 10 and the lowest is 1. The mid value is $(10 + 1)/2 = 5.5$, and 5.5 is the standard value. (2) The mean of all predictions of item i in the training data is the standard value, and the standard values for each item are different. (3) The median of all predictions of item i in the training data is the standard value, and the standard values for each item are different.

Table 3: Average strategy.

	User1	User2	User3	gpred
Item_A	5	1	3	3
Item_B	4	2	3	3

Table 4: Maximum satisfaction.

	User1	User2	User3	gpred
Item_A	5	1	3	5
Item_B	4	2	3	4

Table 5: Calculate disagreement.

	User1	User2	User3	dis
Item_A	5	1	3	4/3
Item_B	4	2	3	4/3

3.3. Group Recommendation Systems Based on External Social-Trust Networks

3.3.1. A Description of the Aggregation Strategy. In group recommendation systems, preference fusion refers to integrating the preferences of group members and preference fusion is also known as the aggregation strategy [11] or the Aggregate Rules [34]; for the unity of terminology, this paper uses the term aggregation strategy. Ricci et al. [11] described 10 aggregation strategies in detail, and having relatively better strategies such as average strategy and average without misery in a series of experiments of another paper.

In average strategy, the group rating for a particular item is computed as the average rating over all individuals.

$$\text{gpred}(G, i) = \frac{1}{|G|}\sum_{u \in G}\text{pred}(u, i), \quad (7)$$

where $|G|$ is the group size and $\text{pred}(u, i)$ is the predicted rating for each user u and every item i. Table 3 shows example of average strategy, and gpred is predicted rating of group.

Maximum satisfaction strategy refers to the greatest rating item in whole group, ignoring the others' lower ratings.

$$\text{gpred}(G, i) = \text{maxpred}(u, i), \quad u \in G, \quad (8)$$

where maxpred refers to max prediction rating in set. Table 4 shows the example about the maximum satisfaction.

Group disagreement with the project [1] $\text{dis}(G, i)$ indicates the degree of difference of users in the group G on the predicted score of the project i.

$$\text{dis}(G, i) = \frac{1}{|G|}\sum_{u \in G}\left[\text{pred}(u, i) - \text{mean}(G, i)\right]^2. \quad (9)$$

$\text{mean}(G, i)$ denotes the mean of the group prediction ratings of item i. Table 5 gives an example of the bifurcation calculation.

3.3.2. GRITrust Algorithm. According to the description of λ and $\text{OAR}(\text{OG}, i)$ on Sections 3.1 and 3.2, the result can

Input: training sets, trust network and G
$D_{\text{train}} = \{\text{user}_i, \text{item}_i, \text{rating}_i\},\ i \in |D_{\text{train}}|$
$\text{Trust\_network} = \{\text{trustor}_j, \text{trustee}_j, \text{value}_j\}$,
$j \in |\text{Trust}|$, and $G = \{u_k, u_m, \ldots, u_n\}$
Output: gpred(G)
Initialization: $\text{pred}(u, i) \Leftarrow$ from D_{train} by CF
repeat
$\text{CGP}(G, i) = \coprod_{\forall u \in G} \text{pred}(u, i)$
$\text{OAR}(\text{OG}, i) = \coprod_{\forall u \in G, v \in OG} t_{uv} \cdot \text{actual}(v, i)$
$\lambda_i \Leftarrow$ from $\text{CGP}(G, i)$ by some methods
if $\text{CGP}(G, i) == \varnothing$ **then**
$\text{gpred}(G, i) = \text{OAR}(\text{OG}, i)$
else $\{\text{OAR}(\text{OG}, i) == \varnothing\}$
$\text{gpred}(G, i) = \text{CGP}(G, i)$
else
$\text{gpred}(G, i) = (1 - \lambda_i) \cdot \text{CGP}(\text{G}, i) + \lambda_i \cdot \text{OAR}(\text{OG}, i)$
end if
until group predictions of all items

Algorithm 1: GRITrust.

Table 6: Actions.

# users	# items	# ratings
1,508	2,071	35,497

Table 7: Trust relationships.

trustor	trustee	trust_value
2	966	1

be computed with those parameters. Note that, if there is no prediction on $\text{CGP}(G, i)$ but on $\text{OAR}(\text{OG}, i)$. The group recommendation systems' result is not empty but $\text{OAR}(\text{OG}, i)$. Above method can ease the cold start problem on GRS. Sometimes, the user has no any ratings, and the system can able to help him to find some better answers.

In this paper, the algorithm named group recommender involve trust (GRITrust) network pseudocode is in Algorithm 1.

4. Experiments

4.1. Experimental Data. In order to verify proposed method all performance experiments were conducted on an open data.

Dataset. We have used the FilmTrust (https://www.librec.net/datasets/filmtrust.zip) [35] ratings dataset for evaluation purposes. The statistics of this dataset are shown in Tables 6 and 7. And rating range is $[0.5, 4]$.

The dataset has been built as follows: firstly, the actual rating of users who have evaluated movies has been divided into 8 to 2, which means the preferences of item for each people; for instance, some users have 10 ratings of different items, training set has 8 ratings, and test set has 2. However some few users have no more than 5 ratings and cannot be split, and we will drop it. Then the training set has 1482 users and test set has 1421 ones, because a bit of users have no rating counts enough. Particularly, only there are 609 users who can trust one or more other ones which means just 41 percent of users would use social network in the training set. Let p be ratio value that is numbers of users on the network to training set population, and the value is 0.41. We can calculate the probability of using social networks as $1 - (1 - p)^n$ via Binomial Distribution when group size is n. And if we can get $n = 2, 3, 4$ then the probability is $0.65, 0.80, 0.87$. Obviously, the value is lower than others when n equals 2. Furthermore, in order to verify the reliability of the proposed method, we would separately discuss the case where the group size is 2.

In this paper, a randomized grouping method is used to conduct the experiment. In the group selection, we noticed that if we experiment with a random group of three people, considering each possibility that brings a large computing cost, for example, when there are 1000 users in dataset, and there are $166,167,000$ randomly selected combinations of 3 people. Obviously, if the numbers and the sizes of groups are appropriately increased, it is impossible to calculate them one by one. For this sake of convenience, we conducted a random sample of studies.

4.2. Evaluation Method Description. In order to evaluate the effectiveness of the proposed method, the root mean square error (RMSE) [7] is used as evaluation method which can assess the quality of the recommended system in terms of accuracy, and the formula is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(\text{gpred}\,(G, i) - \text{actual}\,(G, i)\right)^2}{N}}, \tag{10}$$

where N is the number of items recommended in the group.

Table 8: Explanation.

name	explanation
baseline	without network
GRITrust_dis	λ_{LVD}
GRITrust_mean	λ_{LVTP} α_{mean}
GRITrust_mid	λ_{LVTP} α_{mid}
GRITrust_median	λ_{LVTP} α_{median}

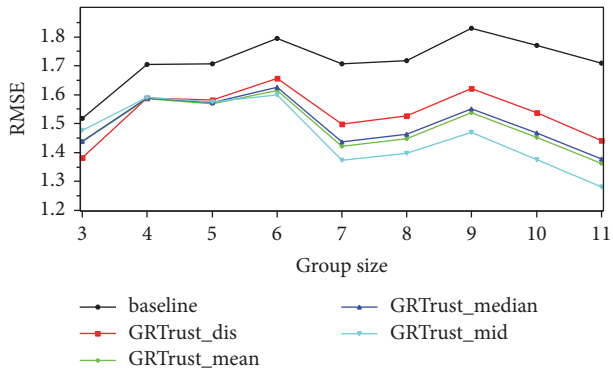


Figure 2: RMSE comparison of different group size by average strategy.

4.3. Randomly Divided into Groups Experiment. The basic idea of the experiment is Collaborative Filtering (CF) algorithm in the training set; each user has not seen item prediction rating, resulting in personalized recommendations. Then randomly select a group and calculate the dynamic adjustment parameters λ, finding the group members trusted in the trust network object, selecting the appropriate aggregation strategy and calculating prediction ratings of item by formula (1). Finally, comparing predictions with the test set, note that the test set also uses the same aggregation strategy.

λ_{LVD} in Table 8 indicates the method of calculating λ by LVD and λ_{LVTP} denotes the method of calculating λ by LVTP. Methods of calculating standard value α are α_{mean} standard value using the true value of the sample; α_{mid} standard value using the middle of the range which can be evaluated; α_{median} standard value using the median of the real sample.

According to group size we have randomly selected 100 times in this experiment, and group size is from 3 to 11. Figure 2 illustrates the performance of baseline, GRITrust_dis, GRITrust_mean, GRITrust_mid, and GRITrust_median with different group size. When group size is 3, GRITrust_dis method is the best than others; it that means that LVP is better than LVTP. In addition, if group size is greater than 5, then LVTP is better than LVP; especially, GRITrust_mid is better than others.

In order to further verify the effectiveness of the proposed method, we carried out the same experiment on the maximum satisfaction strategy. The results of experiment are shown in Figure 3. The GRITrust_dis method is also significantly better than the other methods when group size is 3, but when group size is greater than 5, the error can still be smaller than baseline, and LVTP can get lower RMSE.

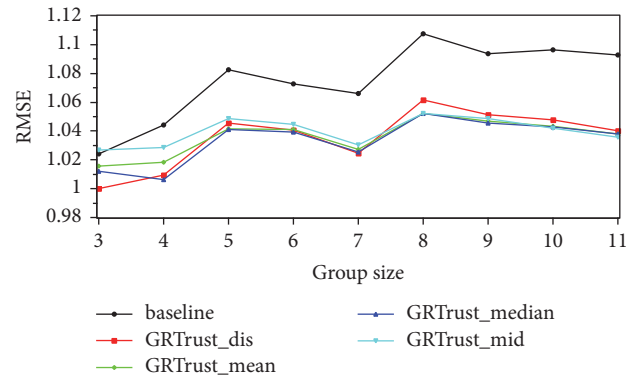


Figure 3: RMSE comparison of different group size by maximum satisfaction strategy.

Through the above two experiments, we show the effectiveness of the proposed method in group recommendation systems, when the large group disagreement is increasing the influence of the social network and vice versa.

4.4. Social Networking Utilization Ratio in Group Recommendation. According to the research, there are three characteristics in group recommendation systems with social network: (1) not all users in social networks, which means that some users do not use social networks; (2) in the current social network, some users do not pay attention to others; (3) some users have little or no rating information on the item. Therefore, we can not guarantee getting the useful information at every time when visiting social networks. The definition of social network utilization is given below.

Definition 1. In group recommendation systems with social network, m is group recommendations, n is social networks visited and used through recommendation system, and we define the ratio of n and m as the social network utilization ratio r_{socail}.

$$r_{\text{socail}} = \frac{n}{m}, \quad r_{\text{socail}} \in [0, 1]. \tag{11}$$

In order to verify the proposed method that is positively correlated with the social network utilization in group recommendation systems with social network, we used the same dataset and method as the random sample experiment except that different utilization ratios of social network were chosen to study the experiment. Among them, $r_{\text{socail}} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$, and the group size is from 3 to 10; for example, G3 indicates that the group size is 3.

In Figure 4, using average strategy, we compared RMSE and social network utilization ratio. The conclusion can be reached that the decline of the Root Mean Square Error has little relationship with the size of the group; however, it is positively correlated with the social network utilization. Figure 5 shows the descent percent by social network utilization ratio. In Figure 6, we evaluated RMSE of 20%, 40%, 60%, 80%, and 100% of social network utilization ratio in different

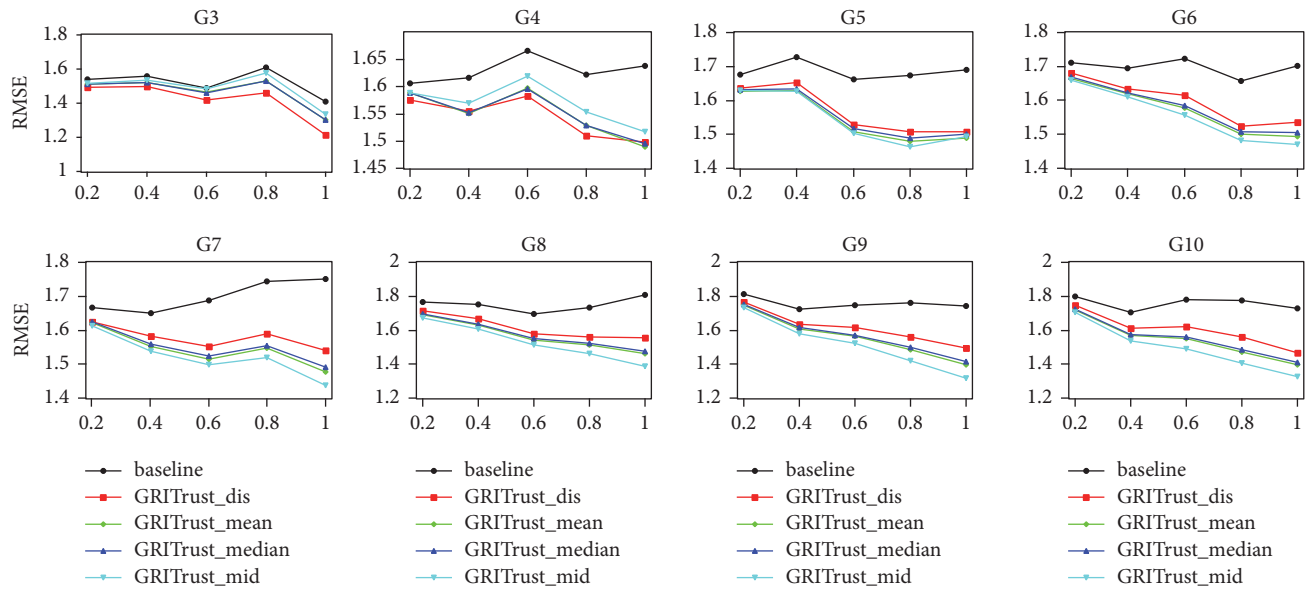


Figure 4: RMSE comparison of different group size and different social network utilization ratio.

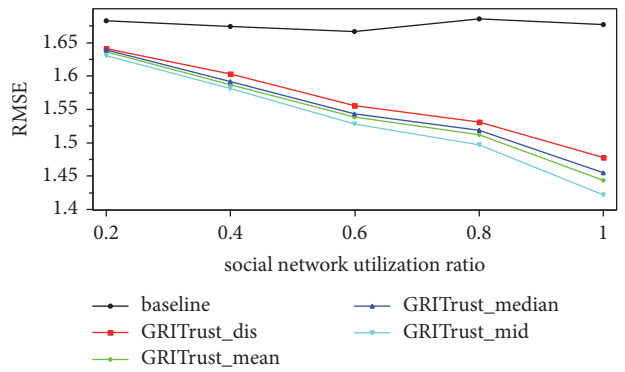


Figure 5: RMSE with social network utilization ratio.

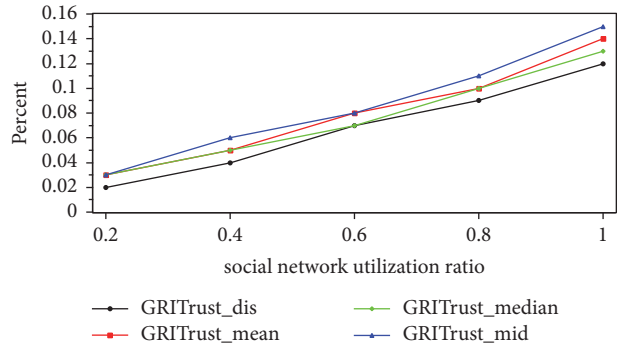


Figure 6: Descent percentage as social network utilization ratio.

group size. The experiment shows that RMSE decreases as the social network utilization ratio increases, and the effect achieved by GRITrust_mid is relatively good, about 4% to 16%.

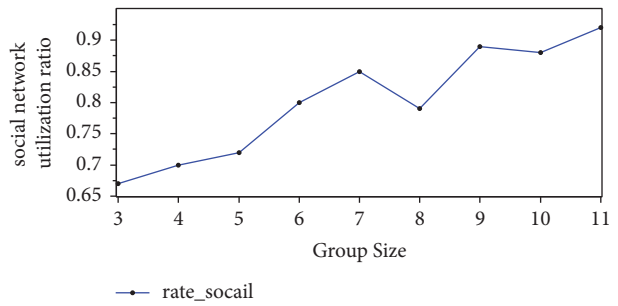


Figure 7: Under random sample.

Figure 7 is the social network utilization in the case of random sampling by different group size; we turn out that the larger the group is, the higher the social network utilization ratio is, and the range is 67%–92% in this dataset. According to Figure 6, RMSE of our proposed method decreases by about 9%–15%.

The above experiment gives the result of using average strategy. In addition, RMSE is reduced as well by maximum satisfaction strategy with different group size in Figure 8. The result of experiment demonstrates the potential of the proposed approach.

Similarly, Figure 9 shows that the RMSE of the proposed method decreases as the social network utilization ratios increase. And GRITrust_median obtained the effect which is relatively good.

In Figure 10 we see that using method λ_{LVTP} is better on average than method λ_{LVD} with different coverage ratios. Among them, the effect of α_{median} in method λ_{LVTP} is the best, and the average drop of root mean square error is the highest, 4.4%.

Through the above experimental verification, in the case of different social network utilization ratio, the higher the

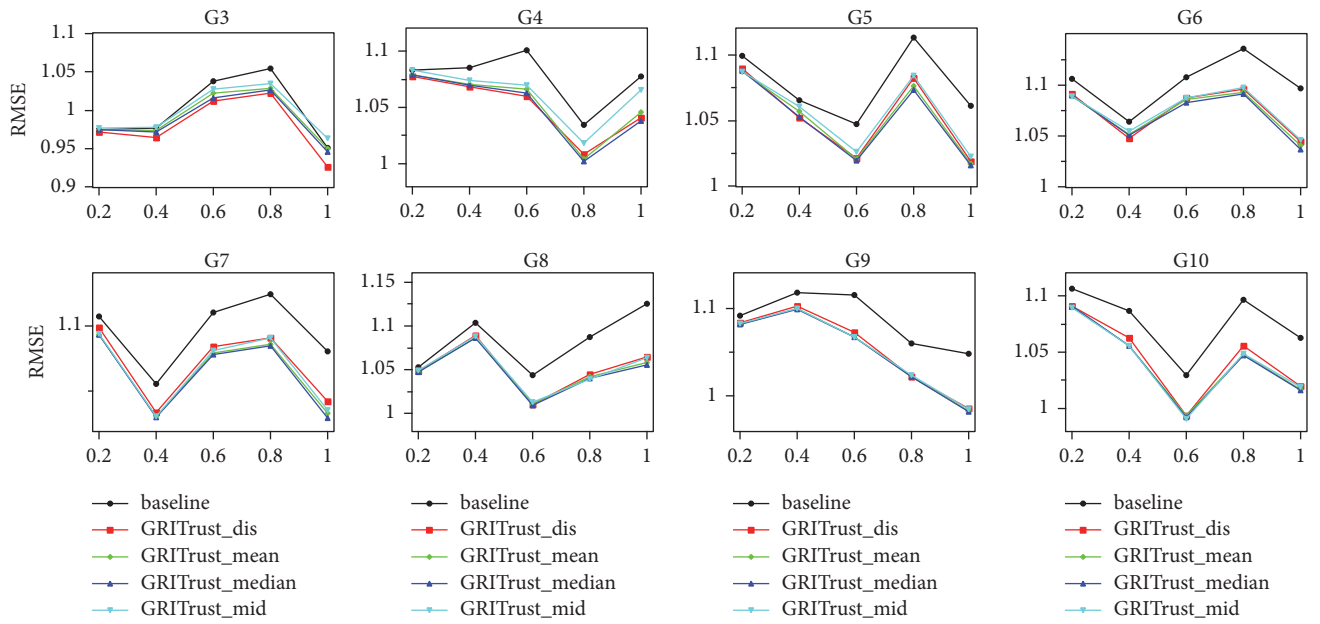


Figure 8: RMSE comparison of different group size and different social network utilization ratio.

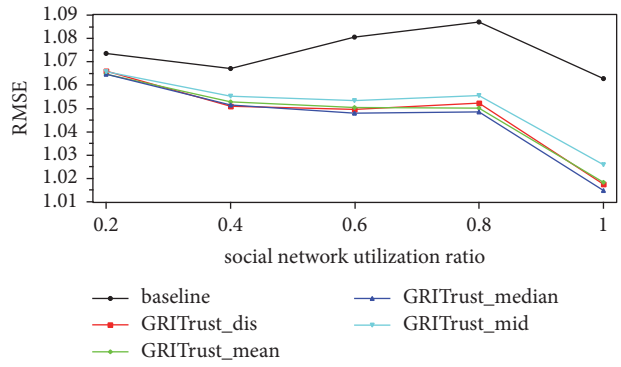


Figure 9: RMSE with utilization of social network.

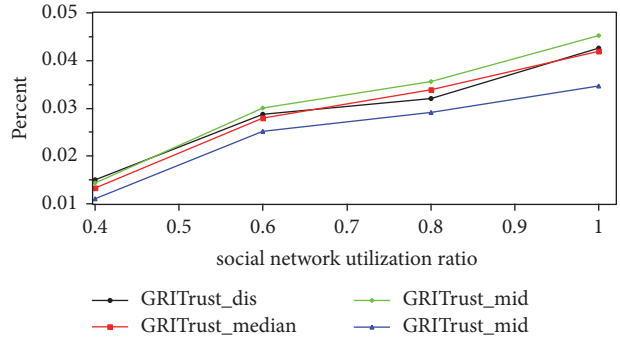


Figure 10: Descent percentage as social network utilization ratio.

utilization rate of social network is, the better the effect of the method proposed in this article is, although the above experiment mainly discusses the case of 3–10 users, ignoring case of group size 2.

Table 9: Social network utilization ratio.

sample size	100	1000	10000
ratio	0.48	0.518	0.5265

4.5. Experiment with Group Size 2. As mentioned earlier, when the number of users in a group is 2, the social network utilization ratio is 0.65. In fact, we can not achieve this value in a limited sample. On the other hand, according to the three aspects mentioned in Section 4.4, it indicates that the social network utilization ratio would not be too high. As a result we conducted a random sampling experiment, the experimental results shown in Table 9; therefore, we specially discussed group size 2.

In this experiment, the group size is 2, and the random sample times are 1000, because the method λ_{LVD} in the group size 2 according to Table 9 shows that we can not calculate the better results, so this experiment would not be considered. The experimental results are opposite in Figures 11 and 5. In the case of 3–10 users, the results are better than the other methods among the methods by λ_{LVTP} and standard value is GRITrust_mid; however, RMSE of this result is higher than baseline as Figure 11 shows which indicates that the effect was not good. While the standard value is GRITrust_median, the RMSE is relatively low, and the effect is better than other standard values.

Similarly in the maximum satisfaction strategy experiment, Figure 12 can get the same conclusion.

According to the above two comparative experiments, when the group size is 2, the method λ_{LVTP} is adopted, and the root mean square error of taking α_{median} is the standard value that is the lowest, and the effect is the best.

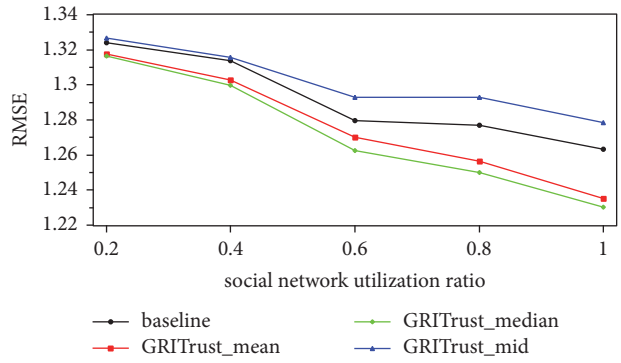


Figure 11: The change curve of RMSE with the social network utilization ratio, average strategy.

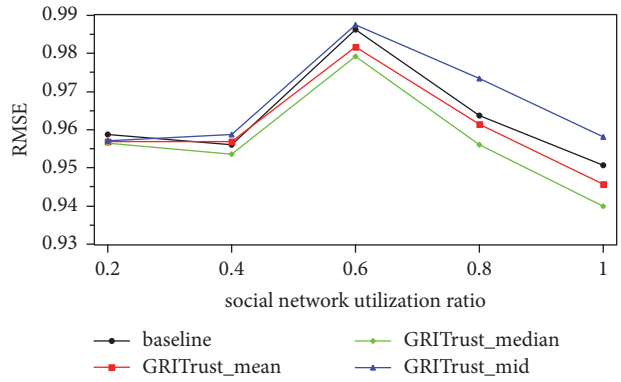


Figure 12: The change curve of RMSE with the social network utilization ratio, maximum satisfaction strategy.

5. Conclusion

This paper introduces the influence of trust network on the group recommendation system. In the traditional group recommendation, when the group preference diverges, the potential intentions of some group members will be ignored. In this paper, we use the trust network relationship in social networks to introduce group members' external real information, through a true evaluation of an item, to amend the group of a forecast of an item, when the group disagreement is small, that is, within the group to achieve the same case, to reduce the social network recommended to the group impact. Thereby dynamically adjusting the impact of social networking factors improves the quality of group recommendations. Through experiments, different aggregation strategies are used to verify the effectiveness of the proposed method in different group sizes. The error of the proposed method does not increase with the increase of the group and will remain at a relatively low level. Based on this, we further discuss the influence of social network utilization on the results of the group recommendation system. Our method shows that, in the group recommendation system, under the same group size, the utilization rate of social network is higher and the root mean square error is lower. In other words, the higher social network utilization ratio is, the better group recommendation is obtained. This shows that, for a new user, as long as the user chooses a few trusted objects, our method can get a good result.

Although some results have been achieved, it does not fully utilize all the valid information in social networks, such as user similarity, user personality, and social status. This paper aims to verify the impact of group members outside other than discussing the relationship between group members, which will be the future need to be discussed. In addition, other domains need to be considered as well except movie; the size of the data is also the direction of efforts.

Data Availability

We have used the FilmTrust ratings dataset at http://www.librec.net/datasets/filmtrust.zip for evaluation purposes. The statistics of this dataset is shown in Tables 6 and 7, and rating range is [0.5, 4].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Science Foundation on of China under the Grant no. 61365010.

References

[1] A. Jameson and B. Smyth, "Recommendation to groups," in *Adaptive Web*, pp. 596–627, 2007.

[2] A. Jameson, "More than the sum of its members: Challenges for group recommender systems," in *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI 2004*, pp. 48–54, ita, May 2004.

[3] M. Kompan and M. Bielikova, "Group recommendations: Survey and perspectives," *Computing and Informatics*, vol. 33, no. 2, pp. 446–476, 2014.

[4] L. Quijano-Sanchez, J. A. Recio-Garcia, and B. Diaz-Agudo, "Personality and Social Trust in Group Recommendations," in *Proceedings of the 2010 22nd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 121–126, Arras, France, October 2010.

[5] L. Quijano-Sanchez, J. A. Recio-Garcia, B. Diaz-Agudo, and G. Jimenez-Diaz, "Social factors in group recommender systems," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 1, article no. 8, 2013.

[6] L. Quijano-Sánchez, B. Díaz-Agudo, and J. A. Recio-García, "Development of a group recommender application in a Social Network," *Knowledge-Based Systems*, vol. 71, pp. 72–85, 2014.

[7] M. Gartrell, X. Xing, Q. Lv et al., "Enhancing group recommendation by incorporating social relationship interactions," in *Proceedings of the 16th ACM International Conference on Supporting Group Work, GROUP'10*, pp. 97–106, usa, November 2010.

[8] L. Quijanosanchez, J. Reciogarcia, and B. Diazagudo, *Group recommendation methods for social network environments*, 2011.

[9] I. A. Christensen and S. Schiaffino, "Social influence in group recommender systems," *Online Information Review*, vol. 38, no. 4, pp. 524–542, 2014.

[10] F. Supan, K. Takanori, F. Goonnapa et al., "Group modeling: Selecting a sequence of television items to suit a group of viewers," *User Modeling and User-Adapted Interaction*, vol. 14, no. 1, pp. 37–85, 2004.

[11] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., *Recommender Systems Handbook*, Springer, 2011.

[12] F. Ortega, J. Bobadilla, A. Hernando, and A. Gutiérrez, "Incorporating group recommendations to recommender systems: Alternatives and performance," *Information Processing & Management*, vol. 49, no. 4, pp. 895–901, 2013.

[13] N. A. Najjar and D. C. Wilson, "Differential neighborhood selection in memory-based group recommender systems," in *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014*, pp. 69–74, usa, May 2014.

[14] J. Kelleher and D. Bridge, *An Accurate and Scalable Collaborative Recommender*, Kluwer Academic Publishers, 2004.

[15] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," in *Proceedings of the 4th ACM Recommender Systems Conference, RecSys 2010*, pp. 119–126, esp, September 2010.

[16] S. Berkovsky and J. Freyne, "Group-based recipe recommendations: analysis of data aggregation strategies," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 111–118, Barcelona, Spain, September 2010.

[17] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.

[18] J. Ben Schafer, F. Dan, J. Herlocker, and S. Sen, *Collaborative Filtering Recommender Systems*, Springer, Berlin, Germany, 2007.

[19] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, pp. 285–295, 2001.

[20] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "IDoctor: personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generation Computer Systems*, vol. 66, pp. 30–35, 2017.

[21] Y. Zhang, "GroRec: a group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[22] Y. Zhang, Z. Tu, and Q. Wang, "TempoRec: Temporal-Topic Based Recommender for Social Network Services," *Mobile Networks and Applications*, vol. 22, pp. 1182–1191, 2017.

[23] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, and L. Peng, "CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies," *Mobile Networks and Applications*, vol. 20, no. 3, pp. 348–355, 2015.

[24] J. F. McCarthy and T. D. Anagnost, "MUSICFX: an arbiter of group preferences for computer supported collaborative workouts," in *Proceedings of the 7th ACM Conference on Computer Supported Cooperative Work (CSCW '98)*, pp. 363–372, November 1998.

[25] A. Crossen, J. Budzik, and K. J. Hammond, "Flytrap: Intelligent group music recommendation," in *Proceedings of the 2002 International Conference on intelligent User Interfaces (IUI 02)*, pp. 184-185, usa, January 2002.

[26] C. Baccigalupo and E. Plaza, "A case-based song scheduler for group customised radio," in *Proceedings of the International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, pp. 433–448, 2007.

[27] H. Lieberman, N. W. Van Dyke, and A. S. Vivacqua, "Let's Browse: a collaborative web browsing agent," in *Proceedings of the 1998 11th Annual ACM Symposium on User Interface Software and Technology, UIST-98*, pp. 65–68, November 1998.

[28] K. McCarthy, M. Salamó, L. Coyle, L. McGinty, B. Smyth, and P. Nixon, "CATS: A synchronous approach to collaborative group recommendation," in *Proceedings of the FLAIRS 2006 - 19th International Florida Artificial Intelligence Research Society Conference*, pp. 86–91, Melbourne Beach, Florida, USA, May 2006.

[29] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso, "Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices," *Applied Artificial Intelligence*, vol. 17, no. 8-9, pp. 687–714, 2003.

[30] J. F. Mccarthy, *Pocket restaurantfinder: A situated recommender system for groups*, 2002.

[31] S. Shin, S.-J. Jang, and S.-P. Lee, "The user-group based recommendation for the diverse multimedia contents in the social network environments," in *Proceedings of the 9th International Conference on Dependable, Autonomic and Secure Computing*, pp. 202–206, December 2011.

[32] J. K. Kim, H. K. Kim, H. Y. Oh, and Y. U. Ryu, "A group recommendation system for online communities," *International Journal of Information Management*, vol. 30, no. 3, pp. 212–219, 2010.

[33] J. E. John, "Thomas-kilmann conflict mode instrument," *Group & Organization Management*, vol. 1, no. 2, pp. 249–251, 1976.

[34] J. S. Dyer and R. K. Sarin, "Group preference aggregation rules based on strength of preference," *Management Science*, vol. 25, no. 9, pp. 822–832 (1980), 1979.

[35] G. Guo, J. Zhang, and N. Yorke-Smith, "A novel bayesian similarity measure for recommender systems," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence, IJCAI 2013*, pp. 2619–2625, chn, August 2013.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 2874509, 13 pages
https://doi.org/10.1155/2018/2874509

Research Article

A Multivariant Stream Analysis Approach to Detect and Mitigate DDoS Attacks in Vehicular Ad Hoc Networks

Raenu Kolandaisamy,[1,2] **Rafidah Md Noor**,[1] **Ismail Ahmedy**,[1] **Iftikhar Ahmad**,[1,3] **Muhammad Reza Z'aba**,[1] **Muhammad Imran**,[4] **and Mohammed Alnuem**[4]

[1]*Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia*
[2]*Faculty of Business & Information Science, UCSI University, Jalan Menara Gading, Kuala Lumpur, Malaysia*
[3]*Department of CS and IT, Mirpur University of Science and Technology (MUST), Mirpur 10250 (AJK), Pakistan*
[4]*College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia*

Correspondence should be addressed to Raenu Kolandaisamy; raenu@ucsiuniversity.edu.my
and Rafidah Md Noor; fidah@um.edu.my

Received 29 December 2017; Accepted 1 April 2018; Published 20 May 2018

Academic Editor: Yin Zhang

Vehicular Ad Hoc Networks (VANETs) are rapidly gaining attention due to the diversity of services that they can potentially offer. However, VANET communication is vulnerable to numerous security threats such as Distributed Denial of Service (DDoS) attacks. Dealing with these attacks in VANET is a challenging problem. Most of the existing DDoS detection techniques suffer from poor accuracy and high computational overhead. To cope with these problems, we present a novel Multivariant Stream Analysis (MVSA) approach. The proposed MVSA approach maintains the multiple stages for detection DDoS attack in network. The Multivariant Stream Analysis gives unique result based on the Vehicle-to-Vehicle communication through Road Side Unit. The approach observes the traffic in different situations and time frames and maintains different rules for various traffic classes in various time windows. The performance of the MVSA is evaluated using an NS2 simulator. Simulation results demonstrate the effectiveness and efficiency of the MVSA regarding detection accuracy and reducing the impact on VANET communication.

1. Introduction

Vehicular Ad Hoc Network (VANET) [1] is a wireless network that allows vehicles to interconnect and communicate with other nearby vehicles, Road Side Units (RSU), or roadside infrastructure. In VANET, each vehicle is considered as a network node which is equipped with an On-Board Unit (OBU) and an Application Unit (AU). The nodes may connect and communicate with each other directly (i.e., Vehicle to Vehicle (V2V)) or through RSUs (i.e., Vehicle to Infrastructure (V2I)) [2–4]. This is primarily for alleviating an Intelligent Transport System (ITS) that aims to provide a wide range of applications and services including safety, nonsafety, and infotainment. In most of these applications, a large number of nodes acquire various services from the network, and the service providing node had a certain capability to handle a specific number of requests. When such requests exceed the capability, the service cannot be guaranteed. On the other hand, the service providing node can accept only a limited amount of data at any point in time, and when it receives a higher payload data packet, it suffers from overload. This high payload data also affects the performance of the network [5, 6]. A VANET architecture and its components are depicted in Figure 1.

The vehicles and RSU act as both transmitters and receivers. The mobility of vehicles is continuous and very fast, especially on highways. Thus, the communication links between vehicles are established only for a short period of time; that is, vehicles are rapidly connecting and disconnecting in the network. This is due to the quickly changing topology. However, mobility of vehicles is predictable as they move on prebuilt highways and roads. Hence the motion pattern of the vehicles can be predicted based on the road topology and layout. Nonetheless, there could be some uncertainty in the movement of vehicles depending upon the

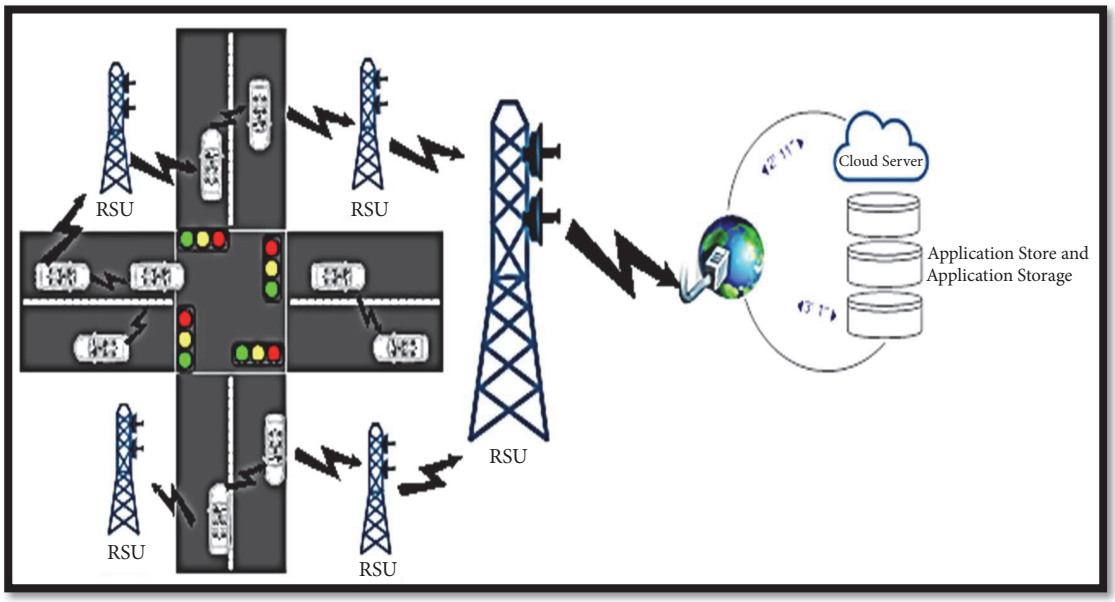


Figure 1: Vehicular Ad Hoc Network architecture.

layout of the road, traffic density, structure of lane, and of course the behavior of the drivers. The nodes in a VANET move at a higher average speed compared to Mobile Ad Hoc Networks (MANETs). The number of nodes in a VANET can be very high on busy highways and very sparse in remote highways. Similarly, at a particular location, traffic is at its busiest during office hours and quiet during midnight hours. Hence any protocol designed should take into consideration these scenarios.

Each vehicular node may acquire a service through various RSU, or the packets might have to travel through several nodes, which makes it vulnerable to Denial of Service (DoS) attacks. In VANET, DoS attacks [6] strive to disrupt the communication channel by flooding it with redundant messages so that legitimate nodes can no longer acquire or use its services. A Distributed Denial of Service (DDoS) attack [6] is more severe as the attack is larger in scale. It involves the participation of multiple nodes across the Internet that the attacker maliciously controls. In a DDoS attack, the attacker may overwhelm the network by using different time slots to send the messages or changing all time slots and messages for different nodes. It is imperative to prevent these types of attacks from crippling the network to allow it to continue its services for safety applications. The objective of this paper is to provide early DDoS detection in VANET environment and make sure the safety of the VANET environment is protected.

1.1. Problem Description. DDoS attack is considered as one of the most severe attacks in VANET. This attack will take down the network to make the service unavailable for the drivers or passengers. This is a vital issue where it may create problem to the drivers on the road and it will particularly be more important if there is life critical information that needs to be transmitted to the drivers. The unavailability of this service or inability to access to it may lead to car accidents [5]. So, this DDoS attack issue cannot be neglected and must be taken seriously. DDoS attack can also occur in any layer of network communication model. The attack will become worse when a DDoS attack which started by more than one perpetrator is executing the attack. This attack is easy to implement and unavoidable for most of the time. In DDoS attack, the attacker controls over the other nodes in network and starts launching attacks from different locations. There are 2 possible scenarios that will happen when a DDoS attack is launched. Figure 2 illustrates DDoS in Vehicle-to-Vehicle communications and Figure 3 illustrates DDoS in Vehicle-to-Infrastructure communications.

(a) Vehicle to Vehicle. Attacker sends messages to victim from different locations or vehicles with the possibility of using different time slot. This attack is to take down the network to make it unavailable for the victim [6].

(b) Vehicle to Infrastructure. Instead of targeting the vehicles, the attacker targets the RSU. The attack will come from different locations and if there are other nodes that wanted to communicate with the RSU, it has already been overloaded. Hence, the service is not available [6].

1.2. Limitations of Existing Approaches. Therefore, there are limited existing solutions for VANETs from DoS and DDoS attacks. The limitations are due to advanced technology and the current threats which are more difficult to prevent. This situation would allow the attackers to detect the ways to intrude into networks. The main limitation of the existing approach is more time is taken to detect the DoS and DDoS

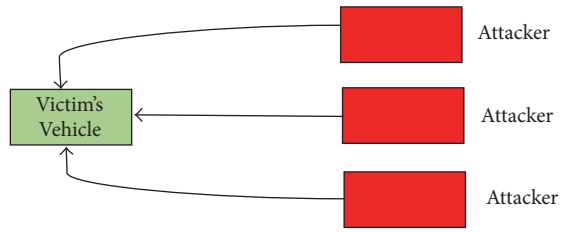


Figure 2: DDoS in Vehicle-to-Vehicle communications.

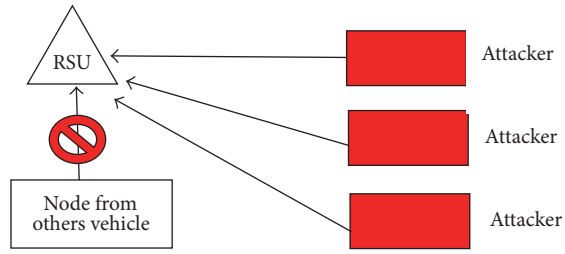


Figure 3: DDoS in Vehicle-to-Infrastructure communications.

attack in VANET environment. Moreover, the existing model has more steps and long algorithm which affects the efficiency and effectiveness [7].

To solve the problems mentioned above, we present a Multivariant Stream Analysis (MVSA) approach to detect and mitigate DoS/DDoS attacks. The proposed approach classifies the traffic into safety and nonsafety applications and RSUs initially maintain structure for V2V communication for generating the traces in the network, to check all node packets. If there is no trace (that means attack), then each packet will be considered as genuine. Uncertainty attack or trace occurs, then it will identify the type of the traffic and compute the multivariant stream factor for each time window. Once identifying the traffic, it will compute the stream weight with the help of traces. Finally, MVSA will classify the effected packet. The performance of the proposed approach is evaluated through simulations. Simulation results demonstrate the efficiency and effectiveness of the proposed approach compared to similar existing approaches regarding various performance metrics such as throughput, detection time, detection accuracy, and ratio.

This paper is organized into six sections: the related works are presented in Section 2. Section 3 describes most common attacks in VANET and discusses safety and nonsafety applications. Section 4 describes the proposed approach. Results and analysis are provided in Section 5. Section 6 presents the conclusion.

2. Related Works

VANET security has been extensively investigated in recent years [1, 7]. However, not much work has been done on identification and mitigation of DoS/DDoS attacks in VANET. Therefore, we specifically focus on this topic in the following.

In [8], the author used Dedicated Short Range Communication (DSRC) and revocation techniques. The detection method is constructed on the offender transfer or sending a message to the target node and then to different locations and may also have a diverse time slot for transferring the message, and the offender will attempt to modify the time slot and the message for different vehicular nodes. However, the main reason for the occurrence is to make the network inaccessible to the victims or vehicle nodes by bringing the entire network down. It has seven channels in DSRC, and the author has created four classes that are sorted based on precedence. Class 1 represents the highest, while class 4 represents the lowest. Nevertheless, some node in the VANET infrastructure will receive a restricted amount of security messages at a specified timestamp, so it is considered as the node that has already been attacked. In this way, it can safeguard itself against any DoS and DDoS attacks.

Another potential method to distinguish DDoS attack is using the Bloom Filter and Traffic Capacity methods [9]. The Bloom Filter is constructed with detection scheme, which is used in providing and protecting against IP spoofing in network addresses. The traffic measurements exposure is based on the detection algorithm, and the algorithm works in three phases. Phase one is in charge of gathering data and the second phase will process the data that has been collected from phase one. If no malicious node is found then the data will be kept in the database. Stage three is the Bloom Filter thru hash function; uncertainly any hateful node was initiated by the second phase, then it generates an alarm and sends the information to the entire nodes in VANET.

The Attacked Packet Detection Algorithm (APDA) [9] and Malicious Node Detection Algorithm (MVND) [10] methods are proposed to detect DoS and DDoS attacks. The APDA method considers time stamp, position, and velocity to detect false or malicious nodes. The method of detecting the malicious node before the verification time will reduce the overhead delay of processing in improving the security in VANET. However, the MVND method is used to detect the

malicious node before the verification time by using a hybrid network. MVND method firstly will allocate the cluster keys by assigning a primary misgiving value to regulate a threshold value by using standard nonconformity and collecting the behavioral data to determine whether the vehicle is abnormal or modified. If it is detected, then it will isolate the vehicle from the network.

The Hybrid Intrusion Detection System (H-IDS) is proposed by the author [11] to detect DDoS attacks. To enhance the overall detection accuracy, the authors combine the anomaly based and the signature-based detection methods. These methods apply for 2 different datasets of the projected scheme to test the H-IDS performance, and the summary of this proposed method provided improved result compared to a system based on the nonhybrid detection. However, two previous works [12, 13] have used the anomaly based method to detect DDoS attacks. The proposed method is not very effective in detecting DDoS; it is because H-IDS method uses two approaches to detect the DDoS attack. However, if we have more than two approaches in one method, it will take some time to complete the process and will affect the detection time.

The Ensemble Based Multifilter Feature Selection method is introduced by the authors in [14] to detect DDoS in cloud computing environments. The proposed method combines the output of 4 filter approaches to attain the best choice which will then evaluate the method with an intrusion detection benchmark dataset and a decision tree classifier. The finding shows that the projected technique can successfully decrease the number of features from 41 to 13 and consumes a top finding rate. The classification accuracy and detection rate are reasonably good compared to other classification techniques. This particular method is used in cloud computing network.

Trilateral trust is based on a defense mechanism compared to DDoS attacks in cloud computing environments [15]. The proposed "trilateral trust mechanism" helps in detecting different kinds of attack groups at different points of time. The direct trust based defense mechanism is for segregating legitimate attack groups from the huge number of incoming requestors. It is a hybrid mechanism of trust that tracks the zero-trust approach initially and eventually supports mutually momentary trust and mutual trust. This combinatorial trust mechanism helps in detecting almost all kinds of overload conditions at a cautionary period. Detecting the high rate of an attack at an earlier moment in time could reduce the traffic impact towards data centers. The results demonstrated that the mechanism proposed is deployable at data centers for resource protection.

Another method is the Queue Limiting Algorithm (QLA) [16], for Defensive VANET from DoS attacks. This proposed scheme works on the safety channels of DSRC to protect the lives of drivers on the road. Classifications have been done for types of application (safety and nonsafety) and DSCR channels. According to the classification, the safety message will trigger first because the safety message is set at a high priority level. In this technique, each vehicle has a restricted size of receiving safety messages. The capacity limit is decided by the proposed algorithm.

Most methods have the problem of poor performance in DDoS detection accuracy, and this paper intends to outline an efficient approach to improve the performance of DDoS detection.

3. Potential Attacks in VANET and Safety and Nonsafety Applications

Interest in the use of VANET is gradually increasing as it improves the safety of passengers. As VANET is used in the open wireless medium, it attracts numerous possible attacks. Hence, the probability of possible attacks is high. The overview of our proposed DDoS attack detection using Multivariant Stream Analysis method is given in this section. The entire detection process consists of three major steps: step 1: preprocessing, step 2: MVSA, and step 3: DDoS mitigation. Figure 2 illustrates the VANET scenario. The web server handles the instruction noted in the RSU through the Internet. The adaptable nature of networks conveys problems associated with security and traffic safety. Network accessibility has been pretentious straight in the situation of DDoS and DoS attacks, wherever the DDoS attack will occur, then the entire network will collapse [17]. The objective of the offender was to initiate problems for authorized users, and as a consequence, services are not accessible, leading to a DoS attack or DDoS attack [1].

(a) Types of Attack. A description of DoS attacks is provided below.

ID Disclosure. ID disclosure is the uniqueness of other vehicular nodes in the intricate infrastructure network and to identify the present position of the target vehicular node. Ultimately, the offender observes the target vehicular node and sends a dangerous virus to the nearby target node. Once attacked, then they will identify the target node ID and the existing location of the target node. These techniques are used by car rental companies to track their cars [18].

Sending False Information. Sending false information is bogus information intentionally directed by a vehicular node to different vehicular nodes in the VANET network to produce confusion which might lead to misunderstanding of the actual condition. Once the false information has been disseminated, most users will leave the road. The attacker can then subsequently use the road for their personal purposes [18].

Timing Attack. In safety applications, the user should receive accurate information or messages on time without any delay; if it is delayed then it will result in a major accident. Time is a very important concern in safety applications. In this attack, the offender will include time slots to generate delays in the message, and the user will obtain the message after the necessary time [19].

Node Impersonation. A node impersonation is when an attacker alters his or her uniqueness to escape being noticed or detected. The attacker will get a message from the initiator

of the message and make some alterations to contents for his/her benefit [19].

Sybil Attack. A Sybil Attack is when a vehicular node directs various messages to different vehicular nodes and every message consists of dissimilar invented source distinctiveness in such a method that the creator does not recognize. The main reason for the attacker is to complicate other vehicular nodes by directing the erroneous messages and to different vehicular nodes consent the road for the profit of the attacker [19].

Denial of Service Attack. In this type of attack, an attacker strives to make the communication channel unavailable for the legitimate vehicular nodes by techniques such as channel jamming. In this case, the affected nodes are unable to send and receive messages [20].

Distributed Denial of Service Attack. The DDoS attacks are produced by DoS attacks [20]. Many offenders launch DoS attacks commencing from dissimilar positions. The offender used altered time slots intended for transferring the messages and time slot of the messages. However, the information may be different from V2V by the attacker. The main reason for the offender is to bring down the entire VANET network in a DoS attack. The circumstance is that the attacker might attack both infrastructure and nodes.

(b) Safety and Nonsafety Applications. As stated earlier, VANET applications can be categorized into safety and nonsafety. The former is more life critical as they are developed to confirm the protection of vehicles and passengers [21, 22]. The latter aims to provide comfort and infotainment to travelers and they can be further divided into pragmatic- and expediency-oriented applications. Table 1 summarizes the different classes of applications and their usage.

4. Multivariant Stream Analysis

In this section, we describe the proposed Multivariant Stream Analysis (MVSA) approach for detection and mitigation of DDoS attacks.

4.1. Preprocessing Stage. In the preprocessing stage, the classification of the safety and nonsafety application traffic will be used. The network trace is maintained by the node which performs DDoS detection. It is just a log of packets received from different source nodes which contain the information of the features considered in this paper. Each packet received will be processed for classification, because the rule is generated at the boot time using the network trace, but if there is no trace, then each packet will be considered as genuine. At the next boot, the detection node will generate the rule. Algorithm 1 discussed the rules. Conversely, the algorithm will compute the rules to perform DDoS attacks detection.

4.2. Multivariant Stream Weight Stage. Multivariant Stream Weight is the second step after the preprocessing step. It is not necessary for the vehicle to read the trace; a single node may be a vehicle which reads the trace and computes the value. The network trace will specify the traffic type and compute the multivariant stream factor. The multivariant stream factor is computed for each time window. By using computed multivariant stream factor, the method computes the multivariant stream weight. Computed stream weight will be used to perform DoS attack detection. Algorithm 2 discusses stream/traffic weight. Conversely, this algorithm will compute the multiattribute stream weight which is used to perform DoS attack detection.

4.3. DDoS Mitigation Stage. DDoS mitigation is the final step in the MVSA approach. In this stage, the node first reads the network trace from neighbor location and preprocesses the logs. The preprocessing algorithm returns the set of rules. As for the received packet stream, the method will compute the multivariant stream weight, by using the rule set generated and stream weight computed, the method will have classified the affected packet. Algorithm 3 discusses the multiattribute similarity measure and stream weight to classify the packet.

Figure 4 illustrates the VANET scenario. The web server handles the instruction nodes in the RSU through the Internet. A main central management station maintains the overall RSUs. The RSU notices the accidents occurring with vehicles and messages are passed through vehicles in a Vehicle-to-Infrastructure (V2I) communication. The V2V denotes the Vehicle-to-Vehicle communication taking place between the vehicles.

Applications of VANET vary in their requirements according to the timely data delivery. The reply time is for the follow-up of accident avoidance in the neighborhood or barrier on the road which tolerates minimum delays for the route optimization models. A minimum delay is acceptable in noncritical delay-tolerant activity mechanisms. The Multivariant Stream Analysis Model and its functional components are shown in Figure 5. Due to the unpredictable nature of the VANET system and high mobility, the detection of DDoS attacks is more challenging [23].

The MVSA method classifies the traffic based on the type of application. Nevertheless, the method maintains various stream classes. The stream class classifies them into two classes: first is safety application traffic and the second is nonsafety application traffic. Conversely, for each class there is a different rule. The rules will be generated according to the number of time windows used, ranging from 1 to 24. As an example, if the class splits time (24) into 1 hour then we will get a 24-time window. The rule will verify the incoming traffic and computes the multivariant stream weight for the incoming packets. Based on computed weight, the method classifies the stream as malicious.

In our model, we are using four parameters. The first is "Payload" which refers to the amount of data present in the packet. The second is "Hop Count," which refers to the number of intermediate nodes a message must have to pass through to reach the destination. The third is "time to live (TTL)," which refers to the lifespan of data in the transmission route or network. However, each data packet has some fixed TTL which is fixed by the MAC layer and the protocol being used. It is also fixed according to the number of hops it has to travel according to the Average Hop

Input: Network Trace Nt.
Output: Ruleset Rs.
Step 1. Start
Step 2. Read network trace Nt.
Step 3. Split trace into different time window.
Step 4. Trace set Ts = $\int_{i=1}^{24}$ Split(Nt, i)
Step 5. For each time window Ti from Ts
Step 6. For each stream class Si
Step 7. Compute average payload Ap = ($\sum$ Ts(Ti, Si).payload)/size($\sum$ Ts(Ti, Si))
Step 8. Compute average hop count Ahc = ($\sum$ Ts(Ti, Si).hop count)/size($\sum$ Ts(Ti, Si))
Step 9. Compute average ttl value Attl = ($\sum$ Ts(Ti, Si).TTL)/size($\sum$ Ts(Ti, Si))
Step 10. Compute average packet frequency Apf = ($\sum$ Ts(Ti, Si))/size($\sum$ Ts(Ti))
Step 11. End
Step 12. Generate Rule Gr = [24 Ahc, Attl, Apf]
Step 13. Add to rule set Rs = $\sum$(Rj $\in$ Rs) $\cup$ Gr
Step 14. End
Step 15. Stop.

Algorithm 1

Input: Network Trace Nt.
Output: MVSW.
Step 1. Start
Step 2. Read network trace Nt.
Step 3. For each time window Ti
Step 4. Compute average payload Ap = ($\sum$ Ts(Ti).payload)/size($\sum$ Ts(Ti))
Step 5. Compute average hop count Ahc = ($\sum$ Ts(Ti).hop count)/size($\sum$ Ts(Ti))
Step 6. Compute average ttl value Attl = ($\sum$ Ts(Ti).TTL)/size($\sum$ Ts(Ti))
Step 7. Compute average packet frequency Apf = ($\sum$ Ts(Ti))/size($\sum$ Ts(Ti))
Step 8. Compute multi-attribute stream factor masv.
Step 9. MASV = $\frac{\text{Ap}}{\text{Apf}} \times \frac{\text{Ahc}}{\text{Attl}}$
Step 10. End
Step 11. Masw = $\frac{\sum \text{MASV}}{24}$
Step 12. Stop

Algorithm 2

Input: Network Trace Nt.
Output: Null.
Step 1. Start
Step 2. Read Network Trace Nt.
Step 3. Rule set Rs = Preprocessing(Nt)
Step 4. Receive incoming packet *P*.
Step 5. Compute multi-attribute stream weight MASW.
Step 6. For each rule Ri from Rule set Rs
Step 7. Compute similarity measure MASM = Dist(Ri.Pl, *P*.Pl)/ $\sum$ Packets receievd in Ti $\times$ Dist(Ri.hc, *P*.hc)/*P*.ttl
Step 8. If MASM < MASW && MASM<>Ri.Features
Step 9. Classify True
Step 10. Else
Step 11. Classify malicious
Step 12. End
Step 13. End
Step 14. Stop.

Algorithm 3

Table 1: Classes of VANET applications and their usage.

Class	Applications	Example usage
Safety oriented	Real-time traffic	(i) RSU stores real-time road traffic data and made it available to vehicles to deal with the problems of traffic jams and avoiding congestion
	Cooperative message transfer	(i) Stopped or slow vehicles to exchange information with other vehicles (ii) Emergency braking to prevent accidents
	Postcrash notification	(i) Vehicles involved in accidents spread warning messages about its location to inform following vehicles
	Road hazard control notification	(i) Disseminating warning messages to other cars about road curves and sudden downhill sections
	Cooperative collision warning	(i) Warning of a driver's capacity on the crash route
	Traffic vigilance	(i) Input: camera installed at RSU (ii) Tool against driving offenses
Pragmatic oriented	Remote vehicle Personalization/diagnostics	(i) Download and install personalized vehicle settings (ii) Uploading of vehicle diagnostics
	Internet access	(i) Through RSU, vehicles can access Internet
	Digital map downloading	(i) Traveler downloads map of region for travel guidance
	Real-time video relay	(i) Traveler watches real-time video
	Value-added advertisements	(i) Online and offline advertisements to attract customers. For example, petrol pumps, 24-hour convenience stores
Expediency Oriented	Route diversions	(i) During road congestion, routes and trips can be planned
	Electronic toll collection	(i) Toll collection via the application. It will help both toll operators and vehicle drivers
	Parking availability	(i) Search for availability of parking slots
	Active prediction	(i) Expect the upcoming terrain
	Environmental benefits	(i) AERIS study program produces and gains environmentally relevant real-time transportation data
	Time utilization	(i) Browse Internet or productive task during traffic jams
	Fuel saving	(i) Vehicle utilizes TOLL system application to pay toll without stopping, saving of fuel approximately 3%

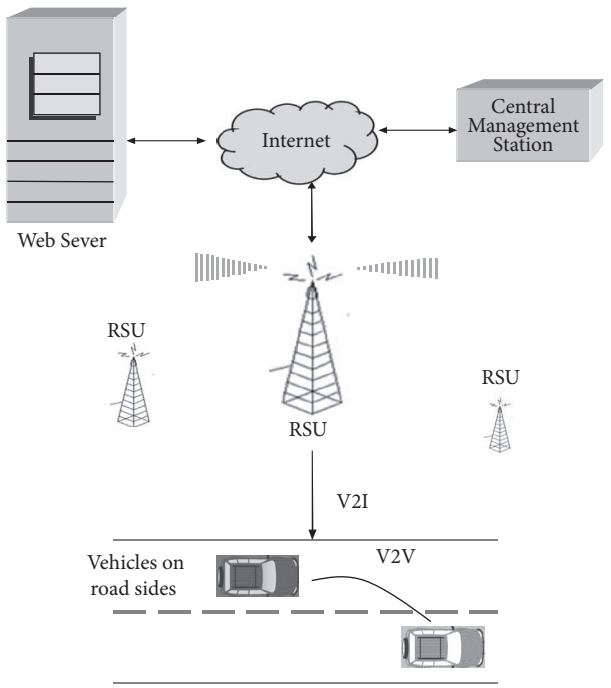


Figure 4: VANET scenario.

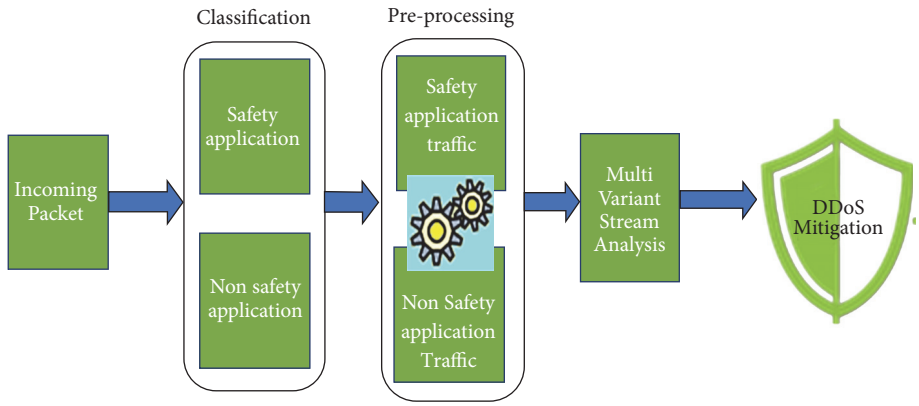


Figure 5: Multivariant Stream Analysis Model.

Table 2: Algorithm rules and its explanation.

Rules	Explanation
Generate rule (Gr) = {Ti, Si, Ap, Ahc, Attl, Apf} Add to rule set (Rs) = $\sum$(Rj $\in$ Rs) $\cup$ Gr	The algorithm will generate the rules according to the Ti, Si, Ap, Ahc, Attl, and Apf. The generated rule will be stored in the set.
$\text{MASV} = \frac{\text{Ap}}{\text{Apf}} \times \frac{\text{Ahc}}{\text{Attl}}$	The average payload is being used it is because various sources share the bandwidth and the bandwidth utilization is depending on the packet frequency as well. Similarly, the TTL value depends on the hop count.
$\text{Masw} = \frac{\sum \text{MASV}}{24}$	The denominator (24) is the entire time value, which is split into the number of the time window. For example, if the class splits time (24) into 1 hour then we will get a 24-time window.
Compute similarity measure MASM = Dist(Ri.Pl, *P*.Pl)/ $\sum$ Packets receievd in Ti $\times$ Dist(Ri.hc, *P*.hc)/*P*.ttl	To compute the similarity, the computed value will be considered. However, computed value for the received packet should fall within the measure of rules that are available for the specific time window. The algorithm must compute the distance between the rules and the features extracted for received packets.
If MASM < MASW && MASM<>Ri.Features	RI. The feature means the feature that is used to detect DDoS attacks. The algorithm has many features in the rule such as time, source, average payload, average TTL, and average hop count. The MASM and MASW are computed according to the mentioned features only. Based on that the decision will be taken.

Count (Ahc). If the packet reaches the destination after the mentioned TTL, then the value is considered as modified or spoofed. So, by counting the TTL value, the chance of being modified can be identified. Nevertheless, if any intermediate node tries to modify or learn the packet features then it will take some time, and it would cross the specified TTL value. Last but not least is the "packet frequency," and the packet frequency is about sending several packets at a particular time. For example, in one minute how many safety application traffic packets have been received and calculate the total number of packets received for safety application.

The incoming packet from V2V and V2I will capture the packet log and send it to the classification stage. In the classification stage, the traffic will identify whether it is safety-oriented or nonsafety-oriented application traffic. Once the traffic is identified, it will go through the preprocessing stage. Once done with the classification process, the preprocessing will generate rules at the boot time using the network trace. If there is no trace, then each packet will be considered genuine. However, the method will read the incoming packet from the classification and split the trace into some classes. One frame is identified for each class, and the method will split the records using traces. The preprocessing will compute the Average Payload (Ap), Average Number of Packets, and Average Hop Count (Ahc). All the three features will compute to generate the rules. The generated rules will be used to perform a DDoS attack. The multivariant stream factor will help to compute each time window. By using computed multivariant stream factors, the method will compute the multivariant stream weight. Computed stream weight will be used to perform DDoS attack detection. Finally, in the DDoS mitigation stage, the rules from preprocessing and stream weight from MVSA will be used to classify the affected packet from the VANET environment. Abbreviations depicts the abbreviation of the algorithm and Table 2 shows the algorithm's rules and its explanations.

5. Results and Discussion

This section describes the simulation setup, performance metrics, baseline approaches, and analysis of results.

Table 3: Simulation configuration.

Parameter	Value
Platform	Ns2
Routing protocol	AODV
Communication range	550 m
Packet size	1000 bytes
Running time	100 Ms (minimum time in network)
RSU	2
Visualization tool	NAM
MAC layer	IEEE 802.11p
Antenna model	Omnidirectional antenna
Traffic type	CBR
Data transmission range	20 Mbps

(a) Simulation Setup. The proposed Multivariant Stream Analysis based DDoS mitigation model has been implemented and evaluated for its efficiency using Network Simulator 2.34. The method has been validated for its efficiency by sometimes maintaining the logs. By using the network trace, the performance of the method for DDoS mitigation was measured. In order to assess the performance, we considered a 4-junction road. In the simulation, the vehicle can initiate a request for its attentive data. However, in the simulation, they were set 5 to 113 vehicles located randomly within the margins. Nevertheless, the vehicle can travel in any direction on the 4-junction road. The time for simulation was executed for 100 Ms. In our simulation, we tested 100 packets and set the simulation time to 100 Ms. Table 3 shows the simulation configuration and parameters for evaluation. However, the mentioned parameters were used in Ns2 to generate simulation to a detected DDoS attack. In this paper we have used AODV routing protocol because our aim is to detect the attack based on routing [1].

There are four junction roads, and they have two lanes in each direction. As shown in Figure 6, there are four crossing junctions through which vehicles may cross each other on the road. In the scenario depicted in the figure, car D is attacked by cars A, C, and E. This is where our proposed model will work to detect the DDoS attack. The result of the simulation is showed in the NAM file, including the trace file routing parameter gained.

(b) Performance Metrics. We measure the proposed model using six different conditions: throughput ratio, packet delay ratio, packet delivery ratio, packet drop ratio, detection accuracy, and detection time [24–27]. The main aim of the performance metrics is to evaluate the performance of MVSA approach to detect the DDoS attack in VANET environments.

Throughput Ratio. Throughput is the factor that is measured based on the number of bytes being sent from the source node towards the destination and the number of bytes being received at the destination at any fraction of the time. Throughput is measured in Kilobits per second (Kbps). For any protocol to prove the efficiency of the protocol, it should achieve higher throughput.

Packet Delay. The packet delivery ratio is the ratio computed between the number of the packets being sent by source node at any point in time and the number of the packets which was received at the destination at the same time window. The same can be measured based on the number of packets received at the destination at any point in time.

Packet Delivery Ratio. Packet delivery ratio depends on the performance of the routing protocol in the VANET network. There is some important parameter to measure the packet delivery ratio, for example, structure of the network, packet size, transmission range, and number of nodes. The packet delivery ratio can be calculated by dividing the number of the packets sent with the number of packets received by the destination. The higher the packet delivery ratio, the better the performance.

Packet Drop Ratio. The packet drop ratio measured using packet did not or never reached the destination from the source network. Normally it will drop in between transmissions.

Detection Accuracy. A detection accuracy is to monitor a network or systems for malicious activity or policy violations. Any detected activity or violation is typically either reported to an administrator or collected centrally using a security information and event management system.

Detection Time. It is measured based on the time at which the packet has been sent from the origin and the time when it has been delivered to the destination. Detection time = (time received − time sent) in milliseconds.

Figure 7 demonstrates the throughput ratio as a function of time when the baseline approaches are compared with the MVSA. The figure clearly shows that MVSA consistently outperforms baseline approaches. This is due to the simplicity of our MVSA method in detecting DDoS attacks, and we did not merge with any other approach. Moreover, the performance of all the approaches improves with the time. This is because the MVSA approach will generate the rules according to the time windows used, ranging from 1 to 24. As an example, if the class splits time (24) into 30 min then we will get a 48-time window. The rules will verify the incoming traffic and compute the MVSW for the incoming packets. It will execute very fast because of the time windows. The throughput ratio of H-IDS is inferior compared to all other approaches because the method is a combination of two approaches. If we combine two approaches, it will take more time to detect a DDoS attack due to increase in the number of steps.

Figure 8 shows the detection accuracy rate as a function of time when the baseline approaches are compared with the MVSA. This figure indicates that MVSA consistently outperforms baseline approaches. This is because our approach will generate the rules from multivariant stream weight, to classify the effected packet accurately and come out with the high accuracy detection. Moreover, the performance of all

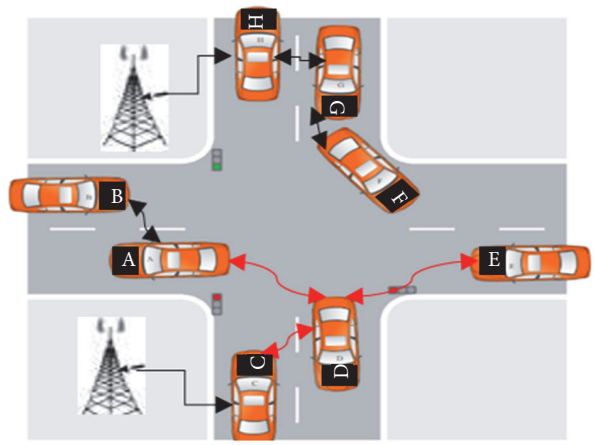


Figure 6: Simulation scenario.

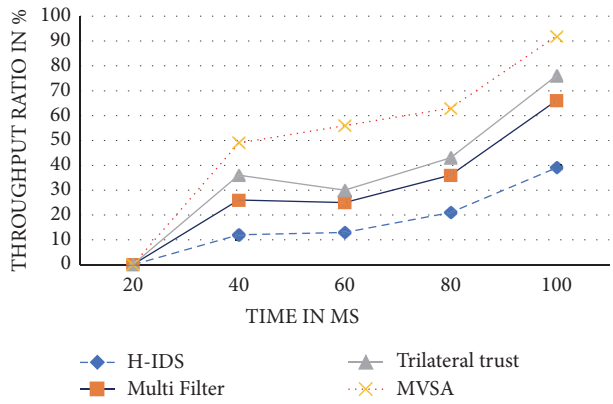


Figure 7: Throughput.

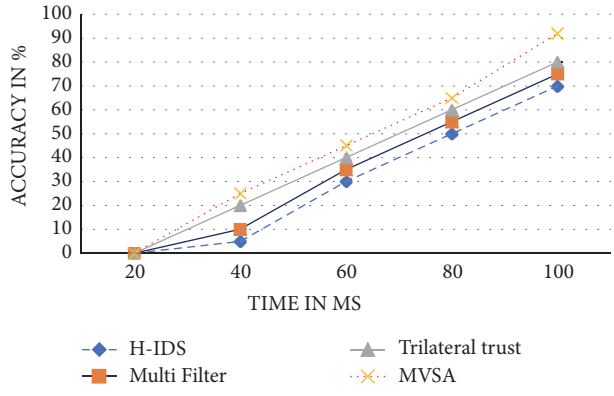


Figure 8: Detection accuracy.

the approaches improves with the time. This is because it takes minimum time to detect in accurate ways. The detection accuracy rate of H-IDS is inferior compared to all other approaches because sometime the vehicle will go far from the neighbor vehicle or RSU.

Figure 9 demonstrates the detection time as a function of time when the baseline approaches are compared with the MVSA. This figure indicates that MVSA consistently outperforms baseline approaches. This is because this approach takes minimum time to detect the DDoS attack compared to another method. Moreover, the performance of all the approaches improves with the time. This is because MVSA approach provides the effective method to detect the attack so that safety application can reach the legitimate user without

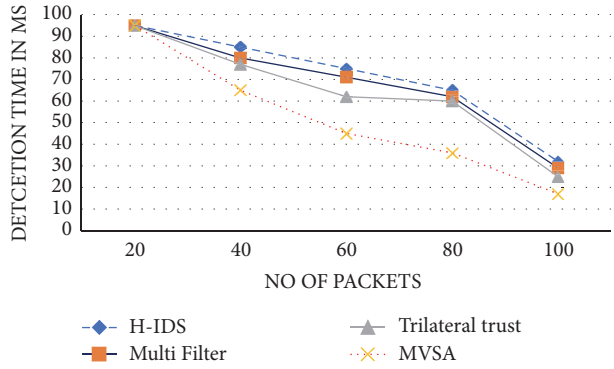


Figure 9: Detection time.

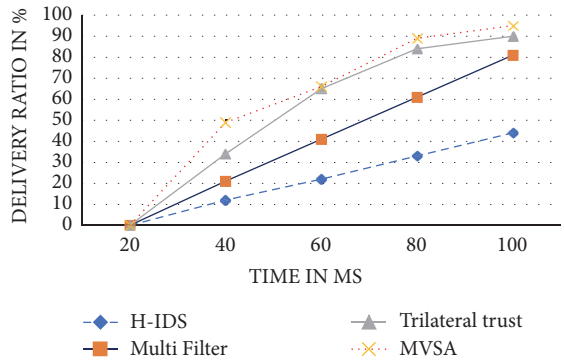


Figure 10: Delivery ratio.

any delay. The detection time of H-IDS is inferior compared to all other approaches because this approach is only focused on throughput and detection accuracy for the network. H-IDS did not focus on detection time, and overall performance is good it will take some time to detect the attack.

Figure 10 demonstrates the packet delivery ratio as a function of time when the baseline approaches are compared with the MVSA. This figure indicates that MVSA consistently outperforms baseline approaches. This is because packet delivery ratio is depending on the performance of the routing protocol in the network. Moreover, the performance of all the approaches improves with the time. This is because if we set more routing protocols then it will take more time to deliver the packet to the destination. Some approach is used for cloud computing network and it will measure the performance of the network. The packet delivery ratio of H-IDS is inferior compared to all other approaches because that approach did not focus on the packet delivery ratio but its more focus on overall throughput, packet delay ration, and detection accuracy.

Figure 11 demonstrates the packet delay ratio as a function of time when the baseline approaches are compared with the MVSA. This figure indicates that MVSA consistently outperforms baseline approaches. This is because of a method that we are using and measurement used based on the stability and performance of the network. Moreover, the performance of all the approaches improves with the time. The packet delay ratio of H-IDS is inferior compared to all other approaches because the approach did not focus on VANET network, its focus on common network. The packet delay is not much different compared with MVSA.

Figure 12 demonstrates the packet drop ratio as a function of time when the baseline approaches are compared with the MVSA. This figure indicates that MVSA consistently outperforms baseline approaches. This is because MVSA approach uses simple method compared with other methods. It is because we spilled rules according to the time windows. Moreover, the performance of all the approaches improves with the time. This is because if we have single process it will reach a destination very fast with less packet drop. If we have more processes it will take more time to process and it will take more time to reach the destination. The packet drop ratios of H-IDS are inferior compared to all other approaches because it focuses on more steps to follow and it also affects the entire packet. Sometimes the vehicle will go far from the neighbor vehicle or the RSU. Its will cause packet drop.

6. Conclusion

In this paper, an efficient Multivariant Stream Analysis (MVSA) approach to detect and mitigate DDoS attacks has been proposed. The vehicle reads the network trace and computes an average measure of payload, time to live, and the frequency for each stream class at different time windows. Four features are measured and computed in the methods to generate the rule set. The rule set is generated, and the features are extracted from the packet received from the user. Nevertheless, the method computes the multivariant stream

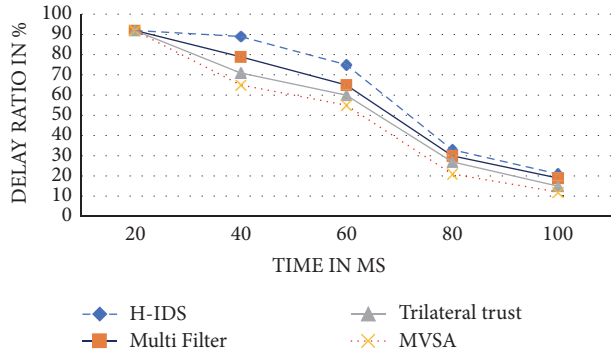


Figure 11: Packet delay.

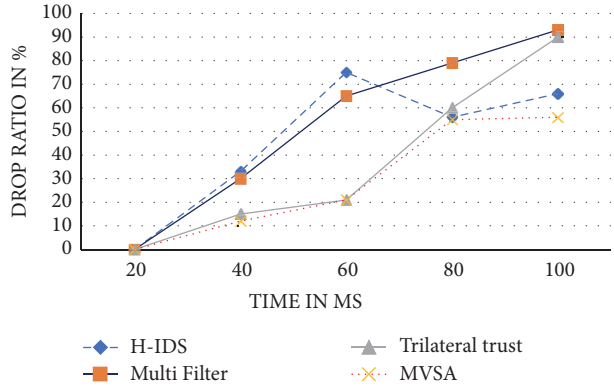


Figure 12: Packet drop ratio.

weight. By using the computed stream weight, the method classifies the packet into either malicious or genuine. The method was shown to be efficient in detecting DDoS attacks in VANET and subsequently reduced the impact on the VANET environment.

Abbreviations

Nt:	Network trace
Gr:	Generate rule
P:	Packet
Rs:	Rule set
Ts:	Trace set
Ap:	Average Payload
Ahc:	Average Hop Count
Apf:	Average Packet Frequency
TTL:	Time to live
Attl:	Average time to live
Ti:	Time window
Si:	Stream class
MVSA:	Multivariant Stream Analysis
MVSW:	Multivariant stream weight
MASV:	Multiattribute stream factor.

Conflicts of Interest

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

Acknowledgments

This work is supported by the Deanship of Scientific Research, King Saud University, through Research Group no. RG-1435-051.

References

[1] H. Hasrouny, A. E. Samhat, C. Bassil, and A. Laouiti, "VANet security challenges and solutions: A survey," *Vehicular Communications*, vol. 7, pp. 7–20, 2017.

[2] I. Yaqoob, I. Ahmad, E. Ahmed, A. Gani, M. Imran, and N. Guizani, "Overcoming the key challenges to establishing vehicular communication: Is SDN the answer?" *IEEE Communications Magazine*, vol. 55, no. 7, pp. 128–135, 2017.

[3] I. Ahmad, R. M. Noor, I. Ali, M. Imran, and A. Vasilakos, "Characterizing the role of vehicular cloud computing in road traffic management," *International Journal of Distributed Sensor Networks*, vol. 13, no. 5, 2017.

[4] I. Ahmad, U. Ashraf, and A. Ghafoor, "A comparative QoS survey of mobile ad hoc network routing protocols," *Journal of the Chinese Institute of Engineers*, vol. 39, no. 5, pp. 585–592, 2016.

[5] L. Li and G. Lee, "DDoS attack detection and wavelets," *Telecommunication Systems*, vol. 28, no. 3-4, pp. 435–451, 2005.

[6] C. Buragohain, M. Jyoti, S. Singh, and D. K., "Anomaly based DDoS Attack Detection," *International Journal of Computer Applications*, vol. 123, no. 17, pp. 35–40, 2015.

[7] S. S. Manvi and S. Tangade, "A survey on authentication schemes in VANETs for secured communication," *Vehicular Communications*, vol. 9, pp. 19–30, 2017.

[8] A. Sinha and S. K. Mishra, "Preventing VANET From DOS & DDoS Attack," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 4, no. 10, 2013.

[9] K. Verma and H. Hasbullah, "Bloom-filter based IP-CHOCK detection scheme for denial of service attacks in VANET," *Security and Communication Networks*, vol. 8, no. 5, pp. 864–878, 2015.

[10] S. A. Ghorsad, P. P. Karde, V. M. Thakare, and R. V. Dharaskar, "DoS attack detection in vehicular ad-hoc network using malicious node detection algorithm," *International Journal of Electronics, Communication and Soft Computing Science & Engineering (IJECSCSE)*, vol. 3, p. 36, 2014.

[11] Ö. Cepheli, S. Büyükçorak, and G. Karabulut Kurt, "Hybrid Intrusion Detection System for DDoS Attacks," *Journal of Electrical and Computer Engineering*, vol. 2016, Article ID 1075648, 8 pages, 2016.

[12] R. Karimazad and A. Faraahi, "An anomaly-based method for DDoS attacks detection using RBF neural networks," in *Proceedings of the International Conference on Network and Electronics Engineering*, 2011.

[13] F. Gong, "Deciphering detection techniques: Part ii anomaly-based intrusion detection," *White Paper, McAfee Security*, vol. 2, p. 1, 2003.

[14] O. Osanaiye, H. Cai, K.-K. R. Choo, A. Dehghantanha, Z. Xu, and M. Dlodlo, "Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, article no. 130, 2016.

[15] N. C. S. N. Iyengar and G. Ganapathy, "Trilateral trust based defense mechanism against DDoS attacks in cloud computing environment," *Cybernetics and Information Technologies*, vol. 15, no. 2, pp. 119–140, 2015.

[16] A. Sinha and S. K. Mishra, "Queue Limiting Algorithm (QLA) for Protecting VANET from Denial of Service (DoS) Attack," *International Journal of Computer Applications*, vol. 86, no. 8, pp. 14–17, 2014.

[17] E. AnkitaThakur and N. Kapoor, *Novel Technique for DDos Attack Isolation in Vanet*, 2017.

[18] K. Verma, "IP-CHOCK reference detection and prevention of denial of service (DoS) attacks in vehicular Ad-Hoc network: Detection and prevention of denial of service (DoS) attacks in vehicular Ad-Hoc network," in *Handbook of Research on Advanced Trends in Microwave and Communication Engineering*, pp. 398–420, IGI Global, 2017.

[19] S. Panjeta, E. K. Aggarwal, and P. Student, "Review paper on different techniques in combination with IDS," *International Journal of Engineering Science*, vol. 11623, 2017.

[20] J. Cheng, X. Tang, and J. Yin, "A change-point DDoS attack detection method based on half interaction anomaly degree," *International Journal of Autonomous and Adaptive Communications Systems*, vol. 10, no. 1, pp. 38–54, 2017.

[21] A. Rasheed, S. Gillani, S. Ajmal, and A. Qayyum, "Vehicular ad hoc network (VANET): A survey, challenges, and applications," *Advances in Intelligent Systems and Computing*, vol. 548, pp. 39–51, 2017.

[22] A. Vaibhav, D. Shukla, S. Das, S. Sahana, and P. Johri, "Security Challenges, Authentication, Application and Trust Models for Vehicular Ad Hoc Network- A Survey," *International Journal of Wireless and Microwave Technologies*, vol. 7, no. 3, pp. 36–48, 2017.

[23] I. Ahmad, I. Ahmad, F. Amin et al., "Towards Intrusion Detection to Secure VANET-Assisted Healthcare Monitoring System," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 6, pp. 1391–1398, 2017.

[24] P. Patel and R. Jhaveri, "A Honeypot Scheme to Detect Selfish Vehicles in Vehicular Ad-hoc Network," in *Computing and Network Sustainability*, vol. 12 of *Lecture Notes in Networks and Systems*, pp. 389–401, Springer, 2017.

[25] V. Saritha, P. V. Krishna, S. Misra, and M. S. Obaidat, "Learning automata based optimized multipath routingusing leapfrog algorithm for VANETs," in *Proceedings of the 2017 IEEE International Conference on Communications, ICC 2017*, IEEE, Paris, France, May 2017.

[26] S. A. Shah, E. Ahmed, M. Imran, and S. Zeadally, "5G for Vehicular Communications," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 111–117, 2018.

[27] I. Yaqoob, E. Ahmed, M. H. U. Rehman et al., "The rise of ransomware and emerging security challenges in the Internet of Things," *Computer Networks*, vol. 129, pp. 444–458, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5913981, 13 pages
https://doi.org/10.1155/2018/5913981

Research Article

On-Demand Mobile Data Collection in Cyber-Physical Systems

Liang He,[1] Linghe Kong,[2] Jun Tao,[3] Jingdong Xu,[4] and Jianping Pan[5]

[1]*University of Colorado, Denver, CO, USA*
[2]*Shanghai Jiaotong University, Shanghai, China*
[3]*Southeast University, Nanjing, Jiangsu, China*
[4]*Nankai University, Tianjin, China*
[5]*University of Victoria, Victoria, BC, Canada*

Correspondence should be addressed to Linghe Kong; linghe.kong@sjtu.edu.cn

Received 10 October 2017; Revised 16 January 2018; Accepted 19 March 2018; Published 30 April 2018

Academic Editor: Yin Zhang

The collection of sensory data is crucial for cyber-physical systems. Employing mobile agents (MAs) to collect data from sensors offers a new dimension to reduce and balance their energy consumption but leads to large data collection latency due to MAs' limited velocity. Most existing research effort focuses on the offline *mobile data collection* (MDC), where the MAs collect data from sensors based on preoptimized tours. However, the efficiency of these offline MDC solutions degrades when the data generation of sensors varies. In this paper, we investigate the on-demand MDC; that is, MAs collect data based on the real-time data collection requests from sensors. Specifically, we construct queuing models to describe the *First-Come-First-Serve*-based MDC with a single MA and multiple MAs, respectively, laying a theoretical foundation. We also use three examples to show how such analysis guides online MDC in practice.

1. Introduction

Collecting data from sensors is a core function of large cyber-physical systems such as wind farm and smart grid [1–4]. Traditional data collection approaches rely on the wireless communications between sensor nodes and the sink, excessively consuming nodes' limited energy supply and leading to their unbalanced energy consumption. Adopting mobile agents (MAs) for data collection, that is, *mobility-assisted data collection* (MDC), reduces and balances the communication loads of nodes (and thus their energy consumption) [5–7]. Also, with MAs' controllable mobility, the communications and networking become possible even in sparse networks via the *store-carry-forward* approach. A real-life MDC example is the NEPTUNE project—a seabed crawler is deployed to collect sensory data from other underwater experiment nodes [8]. However, MDC leads to large data collection latency due to MAs' limited velocity, degrading the realtimeness of the collected data, and may cause data loss due to the buffer overflow at sensor nodes.

Much research effort on *offline* MDC exists in the literature, where the MAs *periodically* collect data from nodes with a preoptimized path [9]. On the other end of the spectrum, *on-demand* MDC—sensor nodes send data collection requests to the MAs when they have data to report and the MAs only visit (and collect data from) such requesting nodes—is a more efficient approach to exploit MAs' limited mobility resource, especially for event-driven systems with diverse data generation among sensors [10, 11]. The challenge in the on-demand MDC, however, is to determine how the MAs should collect data from nodes without a priori information on future data collection demands.

On-demand MDC shows clear queuing behavior. In this paper, we formulate two queuing models to capture the on-demand MDC with the *First-Come-First-Serve* (FCFS) discipline (FCFS is a simple and natural choice to maintain request fairness and is preferred in certain node-centric scenarios.), on the cases where a single MA and multiple MAs are deployed for data collection, respectively, and corresponding analytical results on the data collection performance are

derived. Furthermore, we use three examples to show how the analysis guides the on-demand MDC in practice: (i) how to use multiple MAs? (ii) When to request data collection? (iii) How likely the requests combination—that is, collect data from multiple sensor nodes at the same location—would happen via the wireless communication between the MAs and nodes? The contributions of this paper include the following:

(i) Formulation of an $M/G/1$ queuing model to capture and analytically evaluate the on-demand MDC when a single MA is deployed for data collection (Section 4)

(ii) An $M/G/c$ queuing model for the case when multiple MAs are deployed, based on which the data collection performance is explored via approximation (Section 5)

(iii) Three examples to show how the analysis guides the on-demand MDC in practice (Section 7)

The rest of this paper is organized as follows. The literature on MDC is briefed in Section 2. We formulate the problem in Section 3. The on-demand MDC with a single MA and multiple MAs is investigated in Sections 4 and 5 and evaluated in Section 6. The practical guidance is presented in Section 7, followed by further discussions in Section 8. The paper concludes in Section 9.

2. Related Work

Observing the advantages of MDC over traditional data collection approaches (e.g., via direct communication or multihop forwarding), much effort has been made to explore the *MDC with a single MA* [7, 11, 12]. For example, Sugihara and Gupta investigated the MDC with the objective of minimizing MA's travel distance in [9]. Zhao et al. proposed a three-layer framework for the offline MDC, including the sensor layer, cluster head layer, and MA layer. The MA collects data according to a preoptimized tour with dual antennas [13]. A unified framework for analyzing the MA's mobility and data collection latency was presented and solutions to the involved subproblems were proposed in [14]. An MA-tracking protocol was proposed in [15], in which the routing structure of data collection requests was additively updated with MA's movement. The joint energy replenishment and data collection was investigated in [16], with the consideration of various sources of energy consumption and time-varying nature of energy replenishment. The colocation of MA and wireless charger has been investigated in [17], with the objective of ensuring sustainable and lossless system operation.

Scalability is a critical bottleneck when only a single MA is used and a potential mitigation is to employ multiple MAs for data collection. An early investigation on the scenario of multiple MAs is [18], where the MAs travel along fixed tracks to collect data from nodes with the consideration of load balancing. A motion planning algorithm for the MAs was proposed in [19], which minimizes the number of MAs according to the constraints in distance and time. This work was extended for applications with strict distance/time constraints, and a data-gathering algorithm with multiple MAs was proposed in [20]. More detailed information on MDC is found in [21] and the references therein.

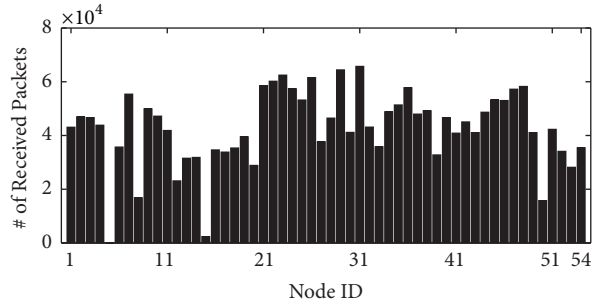


Figure 1: Diverse data generation among sensor nodes (original data from [24]).

Most of these existing efforts focus on the offline MDC, while we tackle the on-demand MDC in this paper. Although similar scenarios have been investigated in [12, 22, 23], our queue-based analytical framework provides detailed insights into the data collection process such as system size, queuing time, and response time, and these analytical insights guide the MDC in practice.

3. Preliminaries

3.1. On-Demand MDC. In many offline MDC solutions, the MAs periodically collect data from sensor nodes based on preoptimized tours [9]. These solutions perform well when nodes generate data at similar paces but degrade dramatically when the data generation at nodes varies—MAs may visit nodes with little or no data to report. Unfortunately, many event-driven systems demonstrate such diverse data generations [10]. For example, Figure 1 plots the number of packets received from 54 sensor nodes in a trace provided by Intel Berkeley Research Lab [24], showing clear diversity among nodes. Targeting on these scenarios with diverse data generations, we investigate the *on-demand* MDC, where the MAs collect data based on real-time demands from sensor nodes in this paper.

3.2. Network Model. We consider the scenario where controllable MAs collect data from stationary sensor nodes randomly deployed in a square sensing field [9] (our model formulation and analysis are also applicable to sensing fields of other shapes, as will be explained in Section 4). Sensor nodes monitor their surrounding environments, store the gathered data in their buffers, and send out data collection requests to MAs when their buffers are to be full [10]. The data collection requests can be delivered to the MAs via existing MA-tracking protocols [15]. Because the typical data relay speed is much faster than the MAs' travel, we assume that the time since a request is sent by a sensor node till it is received by MAs is short and negligible [12]. Note that, instead of using these MA-tracking protocols to upload the sensory data directly, which are normally of much larger volume, only data

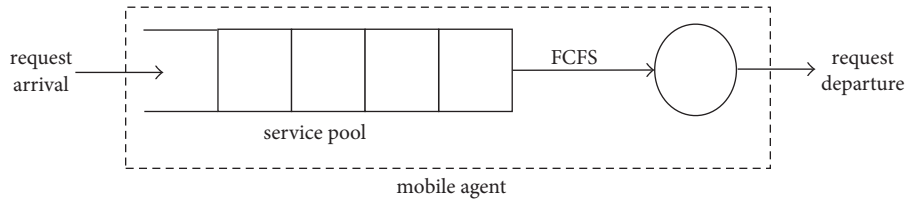


Figure 2: Queuing model when a single MA is deployed for data collection.

collection requests are forwarded to MAs to reduce nodes' communication loads.

MAs maintain a service pool to store the received data collection requests and serve them according to the *First-Come-First-Serve* (FCFS) discipline. By *serving* the request, we mean one of the MAs moves to the corresponding requesting node to collect its data via short-range wireless communications. FCFS, albeit not the optimal solution for on-demand MDC, is a classic scheduling discipline known to be fair for clients [25]. Moreover, the theoretically established data collection performance with FCFS serves as a good baseline for the evaluation of more sophisticated MDC solutions.

Also, it is not necessary for the MA to travel to the exact location of nodes to collect data because of the wireless communications between the MA and the nodes [12, 26]. This way, the MA can potentially collect data from all nodes within its communication range at a single site. The impact of the communication range on MDC relies on both the field size and node density. To establish a theoretic foundation, we do not directly incorporate the communication range into our modeling; however, we investigate and evaluate its impact in Section 7.3.

3.3. Problem Statement. The on-demand MDC is dynamic both temporally and spatially, that is, when a data collection request will be received and where (which sensor node) the request is from. This dynamic property not only shifts our objective from MAs' optimal path planning (as in the *offline* MDC) to the design of efficient real-time service disciplines to select the next request (i.e., the requesting node) to serve (i.e., collect data from) but also makes the MDC hard to capture and thus its performance challenging to evaluate. In this paper, we evaluate the on-demand MDC via a queue-based analytical approach.

4. MDC with a Single MA

We investigate the on-demand MDC in this section and the following ones. Specifically, in this section, an $M/G/1$ queuing model is constructed to capture the MDC when a single MA is deployed.

4.1. Construction of the M/G/1 Model. The on-demand MDC shows clear queuing behavior, inspiring us to capture it with a queuing model—the MA serves as the server and the data collection requests from sensors are treated as the clients (Figure 2). For any queuing system, two fundamental components to be characterized are the client arrival and departure.

4.1.1. Request Arrival. The aggregated request arrival process at the MA is the superposition of n requesting processes of individual sensors, where n is the number of sensors in the system. This way, the request arrival at the MA can be captured by a Poisson process according to *Palm-Khintchine theorem* [27]. This is because (i) for a stable data collection process the number of sensors in the system is large when compared with the number of to-be-served requests at any given time instance, indicating low dependency in their requesting of data collections; (ii) the probability for a sensor to initiate a data collection request at a specific time instance is small. Theoretically, if the client population of a queuing system is large and the probability by which clients arrive at the queue is low at a specific time, the arrival process can be adequately modeled as Poisson [28]. We will further statistically verify this Poisson arrival of requests in Section 6.

Assume that a memory buffer of size B is equipped for each sensor and its asymptotic data generation rates are f_i $(i = 1, 2, \ldots, n)$. The request arrival rate λ can be approximated as

$$\lambda \approx \frac{1}{B}\sum_{i=1}^{n} f_i. \tag{1}$$

Essentially, (1) is a lower bound on the aggregated requests arrival rate because the requesting node would not request again before its data has been collected. Denoting λ^* as the true request arrival rate, we have

$$(\lambda^* - \lambda) \longrightarrow 0^+ \quad \text{as } n \longrightarrow \infty. \tag{2}$$

4.1.2. Request Departure. The MA travels to the requesting node to collect the data therein. Because the data propagation speed is much faster than the MA's travel speed, we simplify our investigation by assuming a negligible data transmission latency. This way, the departure process, or the service time of clients, can be characterized by the time from the service completion of the current request to the time when the MA moves to the next requesting node.

As the previous data collection site is also the starting location when the MA serves the next request, the service time of consecutively served requests seems not to be independent. However, denoting the sequence of service times as $\{t_1, t_2, t_3, \ldots\}$, if we examine only at every second element of

the original process, it is clear that $\{t_1, t_3, \ldots\}$ are independent of each other, and the distribution-ergodic property of this subprocess can be observed [29]. The same is true for subprocess $\{t_2, t_4, \ldots\}$. The distribution-ergodic property still holds if we combine these two subprocesses because their asymptotic behaviors do not change after the combination. A demonstration on this distribution-ergodic property is shown in Figure 3. This means that if we identify the time distribution when the MA travels between consecutively served nodes, we can use it as the service time distribution for the queuing model over a long time period.

From existing results in geometrical probability [11], the distance distribution between two random locations in a unit square is

$$f_{\mathscr{D}}(x) = \begin{cases} 2x\left(\pi - 4x + x^2\right) & 0 \le x \le 1 \\ 2x\left[2\sin^{-1}\left(\frac{1}{x}\right) - 2\sin^{-1}\sqrt{1 - \frac{1}{x^2}} + 4\sqrt{x^2 - 1} - x^2 - 2\right] & 1 \le x \le \sqrt{2} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Thus, with MA travel speed v, the service time distribution can be derived as

$$\begin{aligned} F_{\mathscr{S}}(t) &= P\{\text{service time} < t\} = P\{\text{travel distance} < vt\} \\ &= \begin{cases} \int_0^{vt} 2x\left(\pi - 4x + x^2\right) dx & t \le \frac{1}{v} \\ \int_0^1 2x\left(\pi - 4x + x^2\right) dx + \int_1^{vt} 2x\left[2\sin^{-1}\left(\frac{1}{x}\right) - 2\sin^{-1}\sqrt{1 - \frac{1}{x^2}} - x^2 - 2\right] dx & \frac{1}{v} < t \le \frac{\sqrt{2}}{v} \\ 1 & t > \frac{\sqrt{2}}{v}. \end{cases} \end{aligned} \tag{4}$$

Its expectation, variance, and coefficient of variation are

$$\begin{aligned} \mathbb{E}[\mathscr{S}] &= \mathbb{E}\left[\frac{\mathscr{D}}{v}\right] = \frac{1}{v}\mathbb{E}[\mathscr{D}], \\ \mathbb{V}[\mathscr{S}] &= \mathbb{V}\left[\frac{\mathscr{D}}{v}\right] = \frac{1}{v^2}\mathbb{V}[\mathscr{D}], \\ \mathbb{C}\mathrm{ov}[\mathscr{S}] &= \frac{\sqrt{\mathbb{V}[\mathscr{S}]}}{\mathbb{E}[\mathscr{S}]} = \frac{\sqrt{\mathbb{V}[\mathscr{D}]}}{\mathbb{E}[\mathscr{D}]} = \mathbb{C}\mathrm{ov}[\mathscr{D}]. \end{aligned} \tag{5}$$

After characterizing the request arrival and departure, we can model the on-demand MDC with a single MA as an $M/G/1$ queuing system. Note that the distance distributions between the random locations in other field shapes are also available in the literature [30, 31], which can be used in our model accordingly (e.g., by substituting (3)).

4.2. Analysis Based on the M/G/1 Queuing Model. With the $M/G/1$ queuing model, the data collection latency is equivalently the client response time in the queuing model. We next derive analytical results on the latter to shed light on the former.

4.2.1. System Size Distribution. Denote X_n as the number of requests in the service pool immediately after the departure of a request at time t_n; then

$$X_{n+1} = \begin{cases} X_n - 1 + A_{n+1} & (X_n \ge 1) \\ A_{n+1} & (X_n = 0), \end{cases} \tag{6}$$

where A_{n+1} is the number of new arrivals when serving the $(n+1)$th request. It is clear that A_{n+1} depends only on the service time of the $(n+1)$th request rather than any events that occurred earlier (i.e., the system size at earlier departure points, $X_{n-1}, X_{n-2}, \ldots$). Thus, the embedded discrete-time process $\{X_1, X_2, \ldots\}$ observed at departure times is a *Discrete-Time Markov Chain* (Figure 4) with transition probabilities.

$$p_{ij} = P\{X_{n+1} = j \mid X_n = i\}. \tag{7}$$

Define the probability that i new requests are received when serving a request as

$$\begin{aligned} k_i = P\{A = i\} &= \int_0^\infty P\{A = i \mid S = t\} \times f_{\mathscr{S}}(t)\, dt \\ &= \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^i}{i!} f_{\mathscr{S}}(t)\, dt. \end{aligned} \tag{8}$$

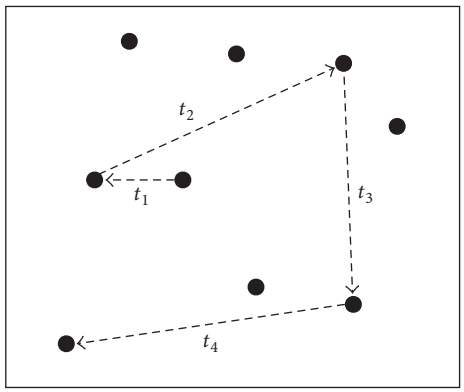


Figure 3: Distribution-ergodic property of the service time.

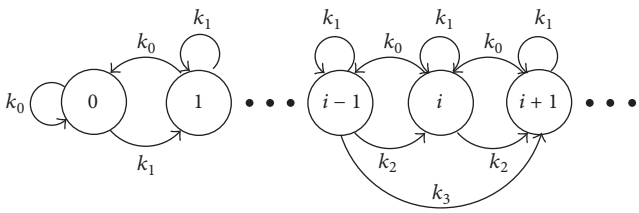


Figure 4: Transition diagram of the Markov Chain.

With k_i, we have the following state transition matrix:

$$\mathbf{P} = \left\{p_{ij}\right\} = \begin{bmatrix} k_0 & k_1 & k_2 & k_3 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ 0 & k_0 & k_1 & k_2 & \cdots \\ 0 & 0 & k_0 & k_1 & \cdots \\ 0 & 0 & 0 & k_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}. \tag{9}$$

Denote $\boldsymbol{\pi}$ as the steady-state system size probabilities at the departure times.

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}, \tag{10}$$

from which we know that

$$\pi_i = \pi_0 k_i + \sum_{j=1}^{i+1} \pi_j k_{i-j+1}, \quad \text{for } i = 0, 1, 2, \ldots. \tag{11}$$

Define the following generating functions:

$$\begin{aligned} \Pi(z) &= \sum_{i=0}^{\infty} \pi_i z^i, \quad |z| \le 1 \\ K(z) &= \sum_{i=0}^{\infty} k_i z^i, \quad |z| \le 1. \end{aligned} \tag{12}$$

Because $\pi_0 = 1 - \rho$, we have

$$\Pi(z) = \frac{(1-\rho)(1-z)K(z)}{K(z) - z}, \tag{13}$$

from which $\boldsymbol{\pi}$ can be derived. Note that $\boldsymbol{\pi}$ are not the same as the steady-state system size probabilities $\{p_n\}$ in general. However, for the $M/G/1$ queue, it has been proven that these two quantities are asymptotically identical [25].

4.2.2. Response Time Distribution. Next we derive the response time distribution based on $\boldsymbol{\pi}$, which consists of two parts: (i) the queuing time since its arrival at the system to its service start and (ii) its service time. With FCFS, for a new request arriving at the queue with n' existing requests, its queuing time is the sum of the service times of these n' requests. By *convolution theorem*, we have

$$q\left(t, n'\right) = \begin{cases} 0 & n' = 0 \\ s(t) & n' = 1 \\ q\left(t, n'-1\right) * s(t) & n' > 1, \end{cases} \tag{14}$$

where $*$ is the convolution operator. This way, the probability distribution of the queuing time can be derived as

$$\begin{aligned} \mathcal{Q}(t) &= \sum_{i=0}^{\infty} \Pr\{i \text{ existing requests}\} \\ &\quad \cdot \Pr\{\text{queuing time} \le t \mid i \text{ existing requests}\} \\ &= \sum_{i=0}^{\infty} \pi_i \int_0^t q(x, i)\, dx, \end{aligned} \tag{15}$$

and its density function is $q(t) = \partial\mathcal{Q}(t)/\partial t$.

Similarly, the response time distribution for a request being received by the MA with a system size n' is

$$r\left(t, n'\right) = \begin{cases} s(t) & n' = 0 \\ q\left(t, n'\right) * s(t) & n' > 1. \end{cases} \tag{16}$$

Its probability distribution can be calculated as

$$\begin{aligned} \mathcal{R}(t) &= \sum_{i=0}^{\infty} \Pr\{i \text{ existing requests}\} \\ &\quad \cdot \Pr\{\text{response time} \le t \mid i \text{ existing requests}\} \\ &= \sum_{i=0}^{\infty} \pi_i \int_0^t r(x, i)\, dx, \end{aligned} \tag{17}$$

and $r(t) = \partial\mathcal{R}(t)/\partial t$.

5. MDC with Multiple MAs

Scalability is a critical bottleneck when only a single MA is deployed for data collection, especially with a large sensing field or with a high node density. Employing multiple MAs to collect data collaboratively is a straightforward mitigation, which we investigate next. Specifically, we consider the scenario where each MA has the full knowledge on the received data collection requests, which can be achieved by the communications among the MAs and the sink, for

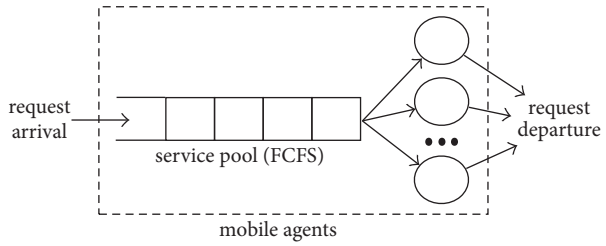


Figure 5: Queuing model when multiple MAs are deployed for data collection.

example, via satellite or cellular communications. Whenever an MA accomplishes its current data collection task, it selects the next-to-be-served request with FCFS.

5.1. Construction of the M/G/c Model. Our approach is to extend the previously constructed $M/G/1$ queuing model to $M/G/c$, where c is the number of MAs (Figure 5). The employment of multiple MAs does not affect sensors' data generation, and thus the request arrival process is the same as in the single MA case. For the request departure, the service time for individual data collection requests is still the same as that with a single MA (i.e., as in (4)), but the aggregated system departure rate will be $c/\mathbb{E}[\mathcal{S}]$.

5.2. Analysis Based on the M/G/c Queuing Model. Although extending the queuing model to the multiple MAs case is straightforward, evaluating an $M/G/c$ queue is analytically intractable. Even when closed-form solutions can be obtained, often they are complicated and require particular probability distributions [32, 33]. Thus, instead of pursuing the exact analytical results on system measures, we use an approximation approach. The basic idea is to combine the analytical results on simple queuing systems such as $M/M/c$ and $M/D/c$ to approximate the measures of the $M/G/c$ queue [34].

5.2.1. Expected Response Time. We first explore the expected data collection latency or, equivalently, the expected response time of the $M/G/c$ queue. A simple two-moment approximation formula with verified accuracy for the mean queuing time in an $M/G/c$ queue can be derived from [35]

$$\mathbb{E}\left[\mathcal{Q}_{M/G/c}\right] \approx \frac{1+\gamma}{2\gamma/\mathbb{E}\left[\mathcal{Q}_{M/M/c}\right] + (1-\gamma)/\mathbb{E}\left[\mathcal{Q}_{M/D/c}\right]}, \quad (18)$$

where $\gamma = \mathbb{C}\mathrm{ov}[\mathcal{S}]^2$.

The above approximation is essentially a weighted combination of $\mathbb{E}[\mathcal{Q}_{M/M/c}]$ and $\mathbb{E}[\mathcal{Q}_{M/D/c}]$. The former can be calculated by

$$\mathbb{E}\left[\mathcal{Q}_{M/M/c}\right] = \frac{(c\rho)^c}{c!c\mu(1-\rho)^2} \cdot \left[\sum_{i=0}^{c-1}\frac{(c\rho)^i}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}\right]^{-1}, \quad (19)$$

and the latter can be obtained by *Crommelin's formula* [36].

$$\mathbb{E}\left[\mathcal{Q}_{M/D/c}\right] = \frac{1}{\mu}\sum_{i=1}^{\infty}\sum_{j=ic+1}^{\infty}\left[\frac{(ic\rho)^{j-1}}{(j-1)!} - \frac{(ic\rho)^j}{\rho j!}\right]e^{-ic\rho}. \quad (20)$$

However, the series in (20) converges slowly especially with high traffic intensity [37]. Again, approximations are adopted to speed up its convergence. An approximation on $\mathbb{E}[\mathcal{Q}_{M/D/c}]$ with simple computation complexity and promising accuracy is presented in [34].

$$\mathbb{E}\left[\mathcal{Q}_{M/D/c}\right] \approx \left[1 + h(\theta)\, g(\rho)\left(1 - e^{-\theta/h(\theta)g(\rho)}\right)\right] \cdot \mathbb{E}\left[\mathcal{Q}_{M/M/c}\right], \quad (21)$$

where

$$\theta = 1 - \frac{2}{c+1}, \quad c \geq 1,$$
$$h(\theta) = \frac{\theta\left[((9+\theta)(1-\theta))^{1/2} - 2\right]}{8(1+\theta)}, \quad (22)$$
$$g(\rho) = \frac{1}{\rho} - 1.$$

Substituting (19) and (21) into (18), we can calculate $\mathbb{E}[\mathcal{Q}_{M/G/c}]$, with which the requests' expected response time in the $M/G/c$ queue can be derived as

$$\mathbb{E}\left[\mathcal{R}_{M/G/c}\right] = \mathbb{E}\left[\mathcal{Q}_{M/G/c}\right] + \mathbb{E}\left[\mathcal{S}\right]. \quad (23)$$

The expected response time is crucial because it not only offers us insights into the asymptotic data collection latency but also helps to obtain the measures on the size of the $M/G/c$ queue, based on which more insights into the MDC can be obtained. Again, denote X as the number of requests either waiting or being served at arbitrary time in the $M/G/c$ queue. Let $\mathcal{P}_i = \mathbb{P}\{X = i\}$ $(i \geq 0)$, and define

$$\alpha = \frac{\mathbb{E}\left[\mathcal{Q}_{M/G/c}\right]}{\mathbb{E}\left[\mathcal{Q}_{M/M/c}\right]}. \quad (24)$$

A geometric-form approximation for the system size probability is proposed in [38].

$$\mathcal{P}_{i,M/G/c} = \begin{cases}\left[\dfrac{(c\rho)^i}{i!}\right]\mathcal{P}_{0,M/M/c} & i = 0,\ldots,c-1\\ (1-\zeta)\zeta^{i-c}\mathcal{U}_{M/M/c} & i \geq c,\end{cases} \quad (25)$$

where

$$\zeta = \frac{\rho\alpha}{1-\rho+\rho\alpha},$$
$$\mathcal{P}_{0,M/M/c} = \left[\sum_{i=0}^{c-1}\frac{(c\rho)^c}{i!} + \frac{(c\rho)^c}{c!(1-\rho)}\right]^{-1}, \quad (26)$$
$$\mathcal{U}_{M/M/c} = \frac{(c\rho_c)^c P_{0,M/M/c}}{c!(1-\rho_c)}$$

is the probability that a newly arriving request has to wait before being served in an $M/M/c$ queue. Note that $\zeta < 1$ if $\rho < 1$ and $\zeta = \rho$ if the service time is exponentially distributed.

We calculate the expected system size based on its approximated distribution as

$$E[\mathscr{L}] = \sum_{i=0}^{\infty} i \cdot \mathscr{P}_{i,M/G/c}, \tag{27}$$

and the probability that a newly arrived request has to wait before being served, that is, the equivalent of $\mathscr{U}_{M/M/c}$ in $M/G/c$ queue, can be calculated as

$$\mathscr{U}_{M/G/c} = 1 - \sum_{i=0}^{c-1} \mathscr{P}_{i,M/G/c}. \tag{28}$$

By *distributional Little's law* [39], the number of customers in the queue has the same distribution as the number of arrivals during the waiting time. Based on this and the above approximation results on the system size distribution, an approximation for the queuing time distribution in the $M/G/c$ queue is proposed in [40].

$$\begin{aligned} \mathscr{Q}_{M/G/c}(t) &\approx 1 - e^{(-c\mu(1-\rho)t)/\alpha} \mathscr{U}_{M/M/c}, \\ q_{M/G/c}(t) &= \frac{\partial Q_{M/G/c}(t)}{\partial t}. \end{aligned} \tag{29}$$

Furthermore, because the response time of a request is the sum of its queuing time and service time, which are independent of each other, by *convolution theorem*, we have

$$r_{M/G/c}(t) = q_{M/G/c}(t) * s(t). \tag{30}$$

6. Performance Evaluations

We verify the model soundness and the analysis accuracy in this section. We consider a system deployed in a square field of size $100 \times 100\,\mathrm{m}^2$. A total number of 100 sensors are randomly deployed unless otherwise specified. The MA velocity is set to 1 m/s based on Power Bot [41]. The simulation is implemented with Matlab. A total number of 10,000 requests are generated and served during each run of the simulation, which is repeated for 50 times.

To deal with the inconvenience of the piecewise distance probability density function in (3), we approximate it by a 10-order polynomial with least squares fitting.

$$\begin{aligned} \tilde{f}(d) = {} & 0.2802d^{10} - 2.0964d^{9} + 2.2349d^{8} \\ & + 24.3629d^{7} - 106.8231d^{6} + 194.4928d^{5} \\ & - 182.8093d^{4} + 91.8223d^{3} - 29.3663d^{2} \\ & + 8.2843d - 0.0402. \end{aligned} \tag{31}$$

6.1. Verifying the Queuing Models. To verify the soundness of the queue-based modeling, we examine the request arrival with an event-driven simulator, where stochastic events occur randomly in the sensing field (note that when events happen in a clustered manner, this actually improves the MDC performance as the MAs' travel distance is reduced. This way, our models capture the worst cases of on-demand MDC, which are important to provide performance guarantees). Sensors within a certain distance (i.e., the sensing range) can detect the event, and corresponding sensory data are generated. The data size for recording each event varies from 10 to 100 B. Events happen independently in both the spatial and temporal domains, and sensor nodes initiate the data collection requests when their buffers become full. We explore the cases where the sensor node buffer size is 4 KB and 8 KB, respectively, and record the interarrival time of data collection requests for comparison with an exponential distribution with the same mean value (the Poisson arrival process indicates an exponentially distributed request interarrival time). Figure 6 indicates that the simulation results match the exponential distribution well, verifying the assumption on Poisson arrival. Furthermore, a larger node buffer results in a smaller request arrival rate, because the sensor nodes can hold the on-board data longer.

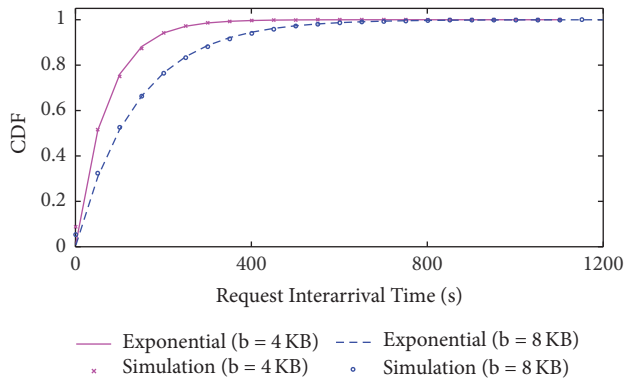


Figure 6: Verify the request arrival.

We further statistically verify the queuing models by validating the Poisson arrival of requests, the independence of request arrivals, and the service time independence. Kolmogorov-Smirnov (K-S) test with a significance level of 5% is used to verify the Poisson arrival. We perform the tests with a different number of sensor nodes (20–100), each with 50 trials. We record the number of trials that reject the Poisson arrival hypothesis. The verification results are listed in the first row of Table 1. The low rejection ratio indicates that the Poisson arrival in our modeling is sound. To evaluate the independence of the request arrival and service time, we record the request interarrival time and service time and calculate their 1-lag autocorrelations. Again, the simulation is repeated for 50 times with 20 to 100 sensor nodes, respectively, and the average absolute values of the autocorrelations are shown in the second and third rows of Table 1. The small correlations of both the request arrival and their service time support our queue-based modeling.

6.2. Single MA. Next, we evaluate our analytical results on the single MA case. Service time distribution is the core

Table 1: Verify the queuing model (with 50 trials).

# of nodes	20	40	60	80	100
# of rejections	0	1	1	3	3
Arrival corr.	0.0463	0.0371	0.0236	0.0183	0.0167
Service corr.	0.0942	0.1023	0.0985	0.0967	0.1000

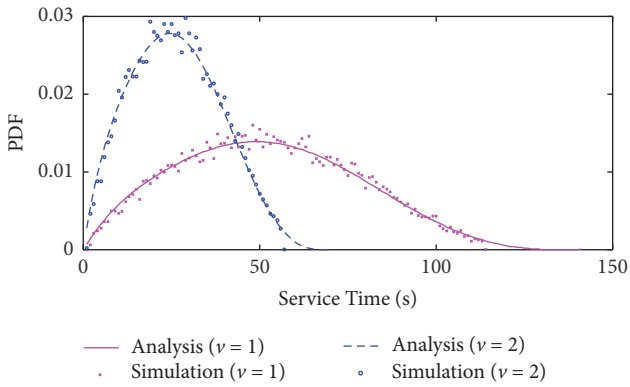


Figure 7: Service time distribution.

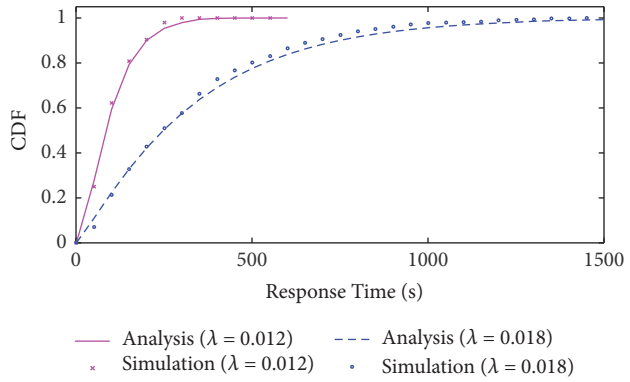


Figure 8: Response time distribution.

component in the queue-based analysis, which is obtained based on results from geometrical probability. We evaluate our analysis on the service time distribution with an MA velocity of 1 m/s and 2 m/s, respectively, and the results are shown in Figure 7. We can see that the analytical results and the simulation match greatly. The service time is significantly reduced after increasing the MA velocity from 1 m/s to 2 m/s, agreeing with (4).

The response time distributions with request arrival rates of 0.012 and 0.018 are shown in Figure 8. Besides the accuracy of the analysis, we can see that the response time of requests, or the data collection latency in our focus, is significantly increased when increasing the request arrival rate. This verifies the potential scalability issue when only one MA is used for data collection.

6.3. Multiple MAs. We evaluate our modeling and analysis results on the multiple MAs case in the following. We explore

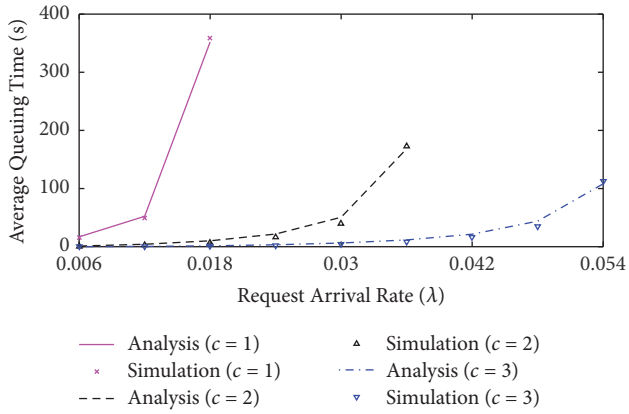


Figure 9: The average queuing time.

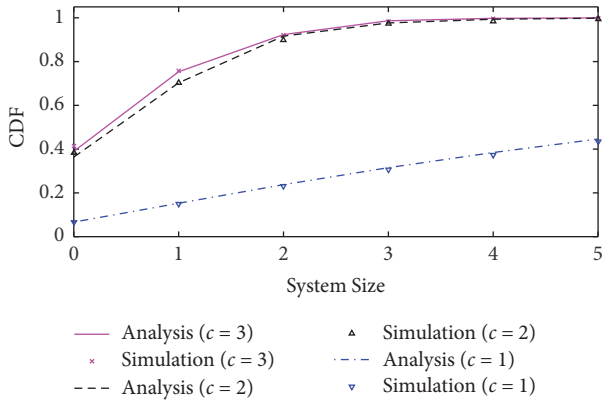


Figure 10: The system size distribution.

the cases with $c = 2$ and $c = 3$, respectively, and also present the results with $c = 1$ for comparison.

The approximation results on the expected queuing time are verified in Figure 9. The effect of deploying more MAs is obvious, especially when the request arrival rate is high. Note that no results for $c = 1$ or $c = 2$ are shown when λ is larger than 0.018 or 0.036, because the further increase of λ will result in a ρ greater than 1, and no steady-state measures can be obtained.

Figure 10 shows the evaluation results of the approximation on the system size distribution with λ of 0.018. Besides the accuracy of the approximation, we can see that increasing c from 1 to 2 can greatly shorten the system size, which in turn reduces the data collection latency. However, the benefit of increasing c further from 2 to 3 is quite limited. This is because the system utilization factor is already small when $c = 2$, and thus further increasing c cannot significantly improve the data collection performance anymore.

The results on the probability for requests to wait before being served are shown in Figure 11. Intuitively, the wait probability increases when the system becomes more heavily occupied, which results when (1) fewer MAs are adopted (c decreases) and (2) the data intensity in the network is higher (λ increases). The verification of this reasoning can be clearly observed from Figure 11.

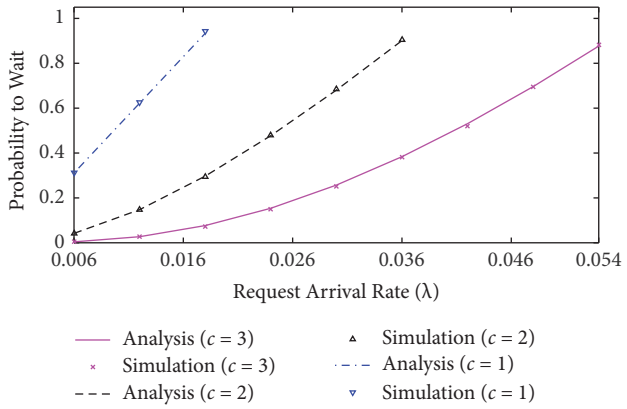


Figure 11: Waiting probability.

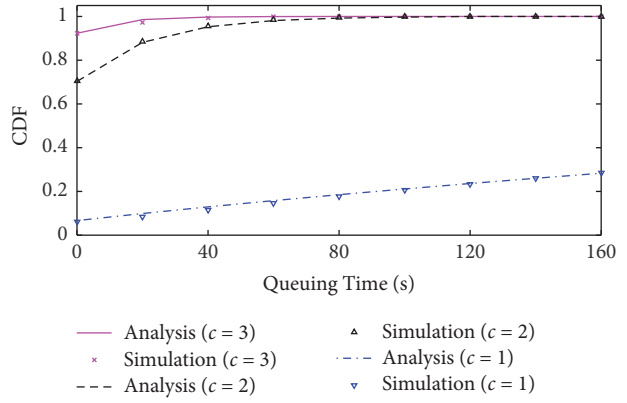


Figure 12: Queuing time distribution.

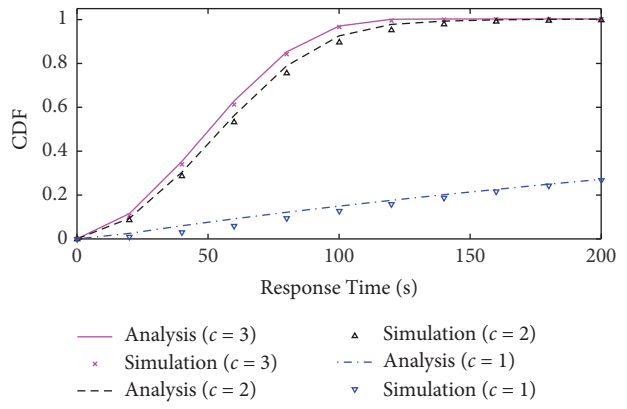


Figure 13: Response time distribution.

The queuing time and response time distributions with λ of 0.018 are shown in Figures 12 and 13, respectively. Besides the analysis accuracy, again we observe that further increasing c from 2 to 3 when $\lambda = 0.018$ cannot significantly reduce the queuing time (response time), agreeing with Figure 10.

7. Practical Guidance

The constructed queuing models not only reveal insights into the on-demand MDC but also guide its practical implementation. We use three examples to show how these models can assist the system implementation in this section.

7.1. How to Adopt Multiple MAs? In the first example, we explore the problem of how to employ multiple MAs for collaborative data collection. In general, two strategies can be used—the MAs can collaboratively collect data from the entire system, referred to as Strategy-I, or the system can be divided into subareas, and each MA is responsible to collect data from one subarea, which is referred to as Strategy-II. These two strategies can be captured by multiserver systems with shared (i.e., all MAs share the knowledge of data collection requests as with Strategy-I) and separate (i.e., each MA is only responsible for a subset of requests that fall in its service queue as with Strategy-II) queues, respectively.

Conventional wisdom says that, all things being equal, a shared queue outperforms separate queues most times [25]. To the best of our knowledge, however, no results on the comparison of the two strategies have been reported yet. Here, we close this gap based on the constructed $M/G/1$ and $M/G/c$ queuing models. Our results reveal that Strategy-II outperforms Strategy-I in both MAs' workloads and requests' response time, contradicting with the conventional wisdom.

Let us consider the case where c MAs are deployed in an $L \times L$ sensing field. For the ease of description, we assume that $c = i^2$ $(i = 1, 2, \ldots)$, which can be relaxed as will be explained later. When Strategy-I is adopted, the data collection performance can be evaluated based on the results in Section 5. Specifically, the utilization factor of individual MAs is

$$\rho_1 = \frac{\lambda}{c}\mathbb{E}[\mathcal{S}] \tag{32}$$

and the expected data collection latency and its distribution can be obtained according to (23) and (30), respectively.

When Strategy-II is adopted, that is, the field is divided into c subareas of size $L/\sqrt{c} \times L/\sqrt{c}$ each, the data collection performance can be evaluated based on the results in Section 4 but in a smaller sensing field. Denoting the requests arrival rate at individual MA and the service time in this case as λ' and $\mathcal{S}'$, the utilization factor of individual MAs is

$$\rho_2 = \lambda'\mathbb{E}\left[\mathcal{S}'\right]. \tag{33}$$

With randomly distributed sensor nodes, it is clear that

$$\begin{aligned} \lambda' &= \frac{\lambda}{c}, \\ \mathbb{E}\left[\mathcal{S}'\right] &= \frac{\mathbb{E}[\mathcal{S}]}{\sqrt{c}}. \end{aligned} \tag{34}$$

Thus, from (32) and (33), we have

$$\frac{\rho_1}{\rho_2} = \frac{\mathbb{E}[\mathcal{S}]}{\mathbb{E}\left[\mathcal{S}'\right]} = \sqrt{c} > 1. \tag{35}$$

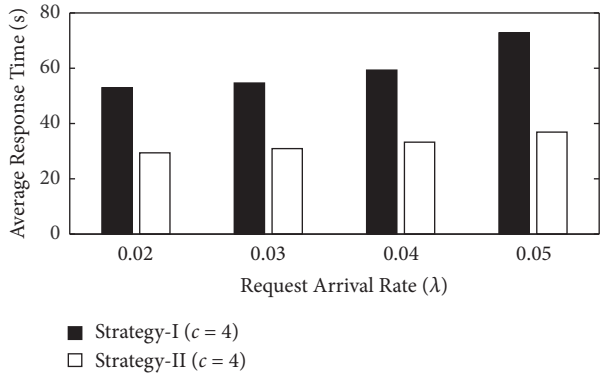


Figure 14: Average response time obtained with the two strategies.

(1) $\theta = 0$, $\Delta = 0.01$;
(2) **while** $p_{\text{overflow}} > 0.01$ and $\theta < 1$ **do**
(3) $\quad \theta = \theta + \Delta; \lambda \dfrac{1}{(1-\theta)B} \displaystyle\sum_{i=1}^{n} f_i$
(4) $\quad$ calculate $R(t)$ with λ;
(5) $\quad p_{\text{overflow}} = 1 - \displaystyle\int_0^{\theta B/f_i} r(t)dt$;
(6) **end while**

Algorithm 1: Find the optimal remaining buffer ratio θ.

This indicates that Strategy-II achieves a lower workload for the MAs with reduced MA travel distance, which dominates the service time in MDC.

Because the comparison on the response time achieved with the two strategies is not so obvious, numerical comparison is performed with a network scale of 100 sensor nodes and $c = 4$. The results are shown in Figure 14, where the aggregated request arrival rate at the MAs varies from 0.02 to 0.05. Again, Strategy-II reduces the response time by 50% when compared with Strategy-I, due to the fact that the service time is dominated by the MAs' travel time.

Although we simplify the above description by assuming that $c = i^2$ ($i = 1, 2, \ldots$), the conclusion on the advantage of Strategy-II holds in more general cases. However, dividing the field into c identical subareas of square shape may not be always feasible, in which cases the distance distributions between random locations in other field shapes can be used [42].

7.2. When to Request Data Collection? The time for sensors to request data collection plays a critical role in the on-demand MDC. Sending the request too early unnecessarily increases the workload of the MAs, which also increases the data collection latency of other requests. On the other hand, a belated request leaves little time for the MAs to complete the data collection before the buffer of the requesting node overflows.

The response time distribution, derived based on the constructed queuing models, helps identify the proper time instant to send out the data collection requests. With a memory buffer B for each sensor node and its respective data generation rate f_i, denote θ ($0 \le \theta \le 1$) as the remaining memory buffer ratio when node sends the data collection request. This way, the aggregated request arrival rate is

$$\lambda \approx \frac{1}{(1-\theta)B} \sum_{i=1}^{n} f_i. \tag{36}$$

Buffer overflow would occur if the response time is larger than $\theta B / f_i$. With the derived response time distribution (i.e., (17) and (30)), the probability for buffer overflow to occur can be calculated with given θ, which in turn allows us to identify the smallest θ (and thus the smallest workloads on MAs) that guarantees a small enough buffer overflow probability (e.g., $p_{\text{overflow}} < 0.01$), as illustrated in Algorithm 1.

7.3. Requests Combination and Preemption. The MAs can potentially collect data from multiple sensor nodes at the same location because of the wireless communication capabilities of both the MAs and sensor nodes, which corresponds to the scenario of batch service in queuing theory and is referred to as *requests combination* here.

Clearly, the probability for requests combination to occur is jointly determined by (i) the communication range R between the MAs and sensors (normalized to the field size) and (ii) the number of requests in the service pool when the new request arrives, that is, the queue length. Specifically, for a new request arriving when L requests are in the queue, it can be combined with at least one of these existing requests with probability

$$P_{\text{combine}}(R, L) = 1 - \left(1 - \int_0^R f_{\mathscr{D}}(x)\, dx\right)^L, \tag{37}$$

where $f_{\mathscr{D}}(x)$ is the distance distribution between two random locations as in (3). This indicates that the effect of requests combination on improving the on-demand MDC will be profound when the communication range is large or when the service queue is long.

Figure 15 shows the effect of requests combination with varying communication ranges. As expected, a larger communication range has a greater effect in reducing the data collection latency, when compared with the noncombination cases. Also, the advantage of requests combination is less significant when the number of MAs increases. This is because the service pool size is reduced when more MAs are adopted for the data collection tasks, and thus the probability for combination to happen is reduced as well.

Requests preemption is another potential way to improve the on-demand MDC—a new request arrival may preempt the service of existing requests if its requesting node is close to the current locations of the MAs. We have analytically explored the possibility and advantage of requests preemption in another work [43].

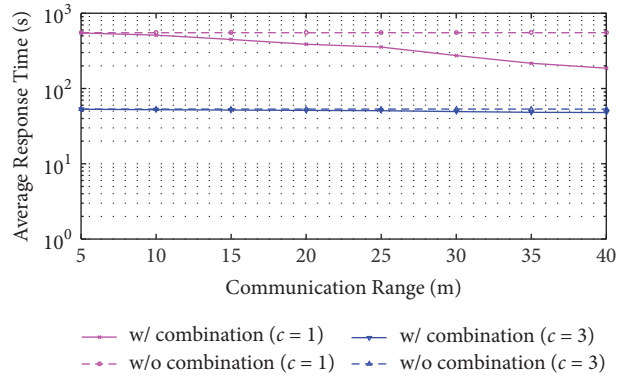


FIGURE 15: Effect of requests combination with communication range.

8. Further Discussions

8.1. System Stability Condition. A necessary and sufficient condition for the data collection process to be stable is $\rho/c < 1$ [12, 22]. This way, from (1), we know that

$$\frac{\rho}{c} = \frac{\lambda \mathbb{E}[\mathcal{S}]}{c} < \frac{\mathbb{E}[\mathcal{D}]}{vcB} \sum_{i=1}^{n} f_i < 1, \tag{38}$$

which implies that the minimum requirement on the travel speed of the MAs is

$$v > \frac{\mathbb{E}[\mathcal{D}]}{cB} \sum_{i=1}^{n} f_i. \tag{39}$$

From (39), we can see that, to provide a stable data collection performance, there is a clear trade-off between the number of required MAs and their capabilities such as travel speed v and memory size B, assisting us in determining the number of needed MAs in practice.

8.2. Insights for Sophisticated Discipline Design. We establish a theoretical foundation on the on-demand MDC when FCFS is adopted, which reveals insights into the design of more sophisticated service disciplines. Through the queue-based analysis, it is clear that the MAs' travel distance between two consecutively served requests is the dominant factor that determines the data collection latency, which should be minimized to achieve a better performance. Inspired by this, in our recent work [11], we have extended these queuing models to investigate the MDC with a greedy service discipline that minimizes the travel distance between two consecutively served requests, and significant asymptotic improvement can be observed. However, the greedy discipline may cause some unfairness among sensor nodes, which has to be addressed to guarantee the worst-case performance for every node. Also note that Petri nets could be another analytical tool to capture such data collection process [44], which we will explore more in the future.

9. Conclusions

In this paper, we have analytically investigated the on-demand MDC in cyber-physical systems. Two queuing models, namely, an $M/G/1$ and an $M/G/c$ model, have been constructed to capture the MDC with a single MA and multiple MAs, respectively. System measures of the queues, for example, the expected values and distributions of queue length, queuing time, and response time have been explored. These queuing models shed light on the impact of different parameters on MDC, and the corresponding analytical results serve as guidelines in the design of more sophisticated data collection solutions. The soundness of the models and the accuracy of the analysis have been verified via extensive simulations.

Through the queue-based analysis, it is clear that the MAs' travel distance between two consecutively served requests is the dominant factor that determines the data collection latency, which should be minimized to achieve a better performance. Inspired by this, in our recent work [11], we have extended these queuing models to investigate the MDC with a greedy service discipline that minimizes the travel distance between two consecutively served requests, and a significant improvement can be observed asymptotically.

Disclosure

A preliminary version of this work was published at IEEE GLOBECOM'11 [45].

Conflicts of Interest

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

Acknowledgments

The work reported in this paper was supported in part by CNS-1739577, National Key Research and Development Program (Grant 2016YFE0100600), NSFC (61672349), the Natural Science Foundation of Jiangsu Province (no. BK20151416), and NSERC.

References

[1] B. Chai, J. Chen, Z. Yang, and Y. Zhang, "Demand response management with multiple utility companies: A two-level game approach," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 722–731, 2014.

[2] M. Dong, K. Ota, L. T. Yang, S. Chang, H. Zhu, and Z. Zhou, "Mobile agent-based energy-aware and user-centric data collection in wireless sensor networks," *Computer Networks*, vol. 74, pp. 58–70, 2014.

[3] G. Li, M. Dong, K. Ota, J. Wu, J. Li, and T. Ye, "Towards QoE named content-centric wireless multimedia sensor networks with mobile sinks," in *Proceedings of the ICC 2017 - 2017 IEEE International Conference on Communications*, pp. 1–6, Paris, France, May 2017.

[4] G. Xie, K. Ota, M. Dong, F. Pan, and A. Liu, "Energy-efficient routing for mobile data collectors in wireless sensor networks with obstacles," *Peer-to-Peer Networking and Applications*, vol. 10, no. 3, pp. 472–483, 2017.

[5] M. Zhao and Y. Yang, "Optimization-based distributed algorithms for mobile data gathering in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 11, no. 10, pp. 1464–1477, 2012.

[6] M. Zhao and Y. Yang, "Bounded relay hop mobile data gathering in wireless sensor networks," *Institute of Electrical and Electronics Engineers. Transactions on Computers*, vol. 61, no. 2, pp. 265–277, 2012.

[7] Y. Gu, Y. S. Ji, J. Li, F. Ren, and B. Zhao, "EMS: efficient mobile sink scheduling in wireless sensor networks," *Ad Hoc Networks*, vol. 11, no. 5, pp. 1556–1570, 2013.

[8] "NEPTUNE Canada," http://www.neptunecanada.ca.

[9] R. Sugihara and R. K. Gupta, "Optimal speed control of mobile node for data collection in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 9, no. 1, pp. 127–139, 2010.

[10] X. Xu, J. Luo, and Q. Zhang, "Delay tolerant event collection in sensor networks with mobile sink," in *Proceedings of the IEEE INFOCOM*, March 2010.

[11] L. He, . Zhe Yang, J. Pan, L. Cai, and J. Xu, "Evaluating service disciplines for mobile elements in wireless ad hoc sensor networks," in *Proceedings of the IEEE INFOCOM 2012 - IEEE Conference on Computer Communications*, pp. 576–584, Orlando, FL, USA, March 2012.

[12] G. D. Celik and E. H. Modiano, "Controlled mobility in stochastic and dynamic wireless networks," *Queueing Systems*, vol. 72, no. 3-4, pp. 251–277, 2012.

[13] M. Zhao, Y. Yang, and C. Wang, "Mobile data gathering with load balanced clustering and dual data uploading in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 4, pp. 770–785, 2015.

[14] Y. Gu, Y. Ji, J. Li, and B. Zhao, "ESWC: efficient scheduling for the mobile sink in wireless sensor networks with delay constraint," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1310–1320, 2013.

[15] Z. Li, Y. Liu, M. Li, J. Wang, and Z. Cao, "Ubiquitous data collection for mobile users in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 312–326, 2013.

[16] S. Guo, C. Wang, and Y. Yang, "Joint mobile data gathering and energy provisioning in wireless rechargeable sensor networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 12, pp. 2836–2852, 2014.

[17] L. Xie, Y. Shi, Y. T. Hou et al., "A Mobile Platform for Wireless Charging and Data Collection in Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1521–1533, 2015.

[18] D. Jea, A. Somasundara, and M. Srivastava, "Multiple controlled mobile elements (data mules) for data collection in sensor networks," in *Proceedings of the 1st IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS '05)*, pp. 244–257, July 2005.

[19] M. Ma and Y. Yang, "Data gathering in wireless sensor networks with mobile collectors," in *Proceedings of the Proceeding of the 22nd IEEE International Parallel and Distributed Processing Symposium (IPDPS '08)*, pp. 1–9, Miami, Fla, USA, April 2008.

[20] M. Ma, Y. Yang, and M. Zhao, "Tour planning for mobile data-gathering mechanisms in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1472–1483, 2013.

[21] Y. Gu, F. Ren, Y. Ji, and J. Li, "The evolution of sink mobility management in wireless sensor networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 507–524, 2016.

[22] E. Altman and H. Levy, "Queueing in space," *Advances in Applied Probability*, vol. 26, no. 4, pp. 1095–1116, 1994.

[23] D. J. Bertsimas and G. V. Ryzin, "A stochastic and dynamic vehicle routing problem in the Euclidean plane," *Operations Research*, vol. 39, no. 4, pp. 601–615, 1991.

[24] "Intel Lab Data," http://www.select.cs.cmu.edu/data/labapp3/index.html.

[25] D. Gross, *Fundamentals of Queueing Theory*, John Wiley & Sons, New Jersey, 4th edition, 2008.

[26] L. He, J. P. Pan, and J. D. Xu, "A progressive approach to reducing data collection latency in wireless sensor networks with mobile elements," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1308–1320, 2013.

[27] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, UK, 1965.

[28] G. Grimmett and D. Stirzaker, *Probability and random processes*, Oxford Science Publications, The Clarendon Press, Oxford University Press, New York, 1982.

[29] C. Bettstetter, H. Hartenstein, and X. Perez-Costa, "Stochastic properties of the random waypoint mobility model," *Wireless Networks*, vol. 10, no. 5, pp. 555–567, 2004.

[30] D. Moltchanov, "Distance distributions in random networks," *Ad Hoc Networks*, vol. 10, no. 6, pp. 1146–1166, 2012.

[31] L. E. Miller, "Distribution of link distances in a wireless network," *Journal of research of the National Institute of Standards and Technology*, vol. 106, no. 2, pp. 401–412, 2001.

[32] B. N. Ma and J. W. Mark, "Approximation of the mean queue length of an *M/G/c* queueing system," *Operations Research*, vol. 43, no. 1, pp. 158–165, 1995.

[33] M. J. Sobel, "Simple inequalities for multiserver queues," *Management Science*, vol. 26, no. 9, pp. 951–956, 1980.

[34] T. Kimura, "Approximations for multi-server queues: system interpolations," *Queueing Systems*, vol. 17, no. 3-4, pp. 347–382, 1994.

[35] T. Kimura, "A two-moment approximation for the mean waiting time in the *GI/G/s* queue," *Management Science*, vol. 32, no. 6, pp. 751–763, 1986.

[36] C. Crommelin, "Delay probability formulae," *Post Office Electrical Engineers Journal*, vol. 26, pp. 266–274, 1934.

[37] G. P. Cosmetatos, "On the Implementation of Page's Approximation for Waiting Times in General Multi-Server Queues," *Journal of the Operational Research Society*, vol. 33, no. 12, pp. 1158-1159, 1982.

[38] T. Kimura, *A transform-free approximation for the queue-length distribution in the finite capacity* M/G/s *queue*, vol. 18 of *Discussion Paper Series A*, Faculty of Economics, Hokkaido University, 1993.

[39] D. Bertsimas and D. Nakazato, "The distributional Little's law and its applications," *Operations Research*, vol. 43, no. 2, pp. 298–310, 1995.

[40] M. H. van Hoorn and H. C. Tijms, "Approximations for the waiting time distribution of the *M/G/c* queue," *Performance Evaluation*, vol. 2, no. 1, pp. 22–28, 1982.

[41] A. Whitbrook, http://robots.mobilerobots.com.

[42] F. Tong, M. Ahmadi, and J. Pan, *Random Distances Associated with Arbitrary Triangles: A Systematic Approach between Two Random Points*, University of Victoria, Victoria, Canada, 2013.

[43] L. He, L. Kong, Y. Gu, J. Pan, and T. Zhu, "Evaluating the On-Demand Mobile Charging in Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 9, pp. 1861–1875, 2015.

[44] G. Liu, "Complexity of the deadlock problem for Petri nets modeling resource allocation systems," *Information Sciences*, vol. 363, pp. 190–197, 2016.

[45] . Liang He, . Jianping Pan, and . Jingdong Xu, "Analysis on Data Collection with Multiple Mobile Elements in Wireless Sensor Networks," in *Proceedings of the 2011 IEEE Global Communications Conference (GLOBECOM 2011)*, pp. 1–5, Houston, TX, USA, December 2011.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 1083575, 12 pages
https://doi.org/10.1155/2018/1083575

Research Article

On the Tradeoff between Performance and Programmability for Software Defined WiFi Networks

Tausif Zahid,[1,2] Xiaojun Hei,[1] Wenqing Cheng,[1] Adeel Ahmad,[3] and Pasha Maruf[3]

[1]*Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*Chenab College of Engineering and Technology, Gujranwala, Pakistan*
[3]*Bahauddin Zakariya University, Multan, Pakistan*

Correspondence should be addressed to Xiaojun Hei; heixj@hust.edu.cn

Received 25 October 2017; Revised 8 December 2017; Accepted 14 January 2018; Published 29 April 2018

Academic Editor: Huimin Lu

WiFi has become one of the major network access networks due to its simple technical implementation and high-bandwidth provisioning. In this paper, we studied software defined WiFi networks (SDWN) against traditional WiFi networks to understand the potential benefits, such as the ability of SDWN to effectively hide the handover delay between access points (AP) of the adoption of the SDWN architecture on WiFi networks and identify representative application scenarios where such SDWN approach could bring additional benefits. This study delineated the performance bottlenecks such as the throughput degradation by around 50% compared with the conventional WiFi networks. In addition, our study also shed some insights into performance optimization issues. All of the performance measurements were conducted on a network testbed consisting of a single basic service set (BSS) and an extended service set (ESS) managed by a single SDN controller deployed with various laboratory settings. Our evaluation included the throughput performance under different traffic loads with different number of nodes and packet sizes for both TCP and UDP traffic flows. Handover delays were measured during the roaming phase between different APs against the traditional WiFi networks. Our results have demonstrated the tradeoff between performance and programmability of software defined APs.

1. Introduction

Many emerging Internet services have been reshaping our daily lives, which are running on wireless portable devices including mobiles and tablets. These services commonly use WiFi networks for Internet access. Studies have shown that in near future most of the IP traffic will be carried wirelessly so there exists a pushing need to increase network capacity and improve its efficiency for end-users [1]. For the streaming applications such as voice over IP, users require Internet services with quality-of-service. Providing network services to different users at the same time may cause hindrance and jitter in the traffic load. The increasing numbers of users and traffic flows everyday suggest that the traditional WiFi networks should be renovated to meet soar demands. The operational cost and traditional infrastructure of the WiFi networks have slowed down this innovation process [2], for example, the radio network resource abstraction and allocation lacks of controlling knots due to the random access on broadcast wireless medium and the dynamic channel conditions. Nowadays, there have been many intelligent devices emerging that have self-adaption awareness which also induce security challenges to the networks. In addition to deliver real-time data, the networks need to be equipped with advanced programmability features to manage these potentially hostile devices. Moreover, the network should provide measurable, manageable, and controllable interfaces to network applications at the upper layer [3].

The SDN paves a new approach of network management by partitioning the control plane and the forwarding plane. In terms of enabling programmability, ability to control network traffic and devices, SDN networks are more flexible than the traditional ones in handling constraints such as channel switching, unbalanced traffic load, and handover. The separation of the control plane and the forwarding plane not only provides flexible management but also provides the centralized control for the whole network. SDN has been accepted as a unique architecture for wired infrastructure, by

providing faster deployment of new services and applications, by enabling novel features such as virtualization. OpenFlow becomes a common south-bound protocol for the SDN deployment [4]. Software defined wireless networking is a natural extension of SDN for wireless networks, which has been proposed in networking research and industry communities in that such a separate controller can control wireless devices in a unified way.

WiFi networks have been shifting from local and independent framework to substantial public infrastructures [5, 6]. It becomes a challenge to provide services to a large-scale network with adequate coverage, low delay, and minimum disruption. A single AP can cover a radius of about 200 to 300 meters in outdoor environments, while in indoor scenarios a single AP can cover about 50 meters; hence, frequent handovers may occur due to limited sizes of hot spots. The distance between APs and end-users has a major impact on bandwidth. A user who is connected with a longer distance to its AP can only receive around 10 to 50 percent of the network bandwidth as compared to the user connected to adjacent AP. It is very important to address this persistent handover problem. In a traditional architecture, it is difficult to configure all the devices in case of small changes in network policy.

Figure 1 illustrates an SDN-based WiFi network architecture, in which multiple network management policies can be controlled under a centralized control. The devices are connected with SDN-based APs, under the management of a single controller which have a global view of networks. In a software defined WiFi network, there is no need to instrument various WiFi protocols on the APs; instead, all the packet forwarding decisions are determined by a centralized controller. By controlling the whole network in form of programmable entities, SDWN offers flexible environments for the management and performance improvement of infrastructure by deploying new services more conveniently.

In this paper, we study the tradeoff between performance and programmability for software defined WiFi networks against the traditional WiFi networks. We instrument an experimental network testbed to conduct performance comparison. This testbed contains a java-based SDN controller, APs and 24 clients for generating real network traffic. This testbed is configured with single and multiple BSS testing scenarios. The performance is measured with TCP and UDP based packets and the evaluation establishes the tradeoff between network performance and control flexibility in SDWN. We focus on a case study to examine the mobility issues in this instrumented SDWN testbed. In SDWN, there are many APs which are managed by such a central controller. This controller serves as the network brain accessible for all the APs. The SDN controller coordinates APs so that the roaming clients are able to maintain network connectivity from one AP to another. Our experiments consider two typical scenarios. In the first scenario, the same channel is used, while different channels are used in the second scenario. The comparison focus on the handover delay in two typical scenarios and the throughput is measured for a SDWN prototype, namely, Odin-V2, in comparison with the conventional WiFi network.

The rest of the paper is organized as follows. In Section 2, we provide a background review on software defined wireless networks. In Section 3, we describe the framework of an experimental SDWN network testbed. Then, we report various performance evaluation results in Section 4. Finally, we conclude the paper in Section 5.

2. Background

With rapid deployment of new Internet based applications [7–9], WiFi has become the most adopted network interface in many portable devices. WiFi services are pervasively available these days, but it is challenging to provide high-speed constant connectivity for many users. Many gaming and voice applications demand continuous network connectivity without any freezing or delay. Mobility is an essential and major issue in cell-based wireless networks. In traditional WiFi network architectures, certain handovers occur due to frequent mobility of a client when he changes his location during connection to one AP. After changing his location, if the client obtains better signal strength from another access point as compared to previously connected AP, this client may establish a new association to this new AP. In a large-scale environment of traditional WiFi networks, this client faces such a frequent handover problem which takes time due to the exchange of management frames between the client and APs. Frequent handover is one of the serious problems in traditional WiFi networks and it is not specified in the 802.11 protocol family. Handover decisions are usually made by vendor specific protocols which take decisions based on signal power, signal to noise ratio, and other criteria. SDN provides advanced network services with flexible and economical hardware, and it is based on a unified way to manage the network [10]. SDN has now been extended for WiFi and cellular networks. The extension of SDN for WLAN has been an active research because many research problems are still open to be addressed including performance and practical deployment of SDWN. OpenFlow is a typical south-bound protocol for a controller to manage network devices, while it is yet to provide the support for 802.11 protocols.

Most available SDN simulators only support wired network infrastructure. It is challenging to study the performance of wireless network because channel interference and other aspects are difficult to replicate in a simulation environment. Odin [11] is a prototype system towards a real deployment of a SDWN. Our testbed heavily utilizes the Odin architecture with necessary module upgrades and various modifications. The virtual access point (VAP) provides management competence and virtualizes the association structure of AP, which empower the administrator to program the network and deploy the WiFi infrastructure. A typical Odin infrastructure contains a single controller and multiple APs. The controller and APs communicate with each other using TCP connections. This deployment has the advantage that no modification is required on the end systems and there is no need of connection reestablishment in case of handover. The management process running on the controller will migrate the connection from one VAP to another

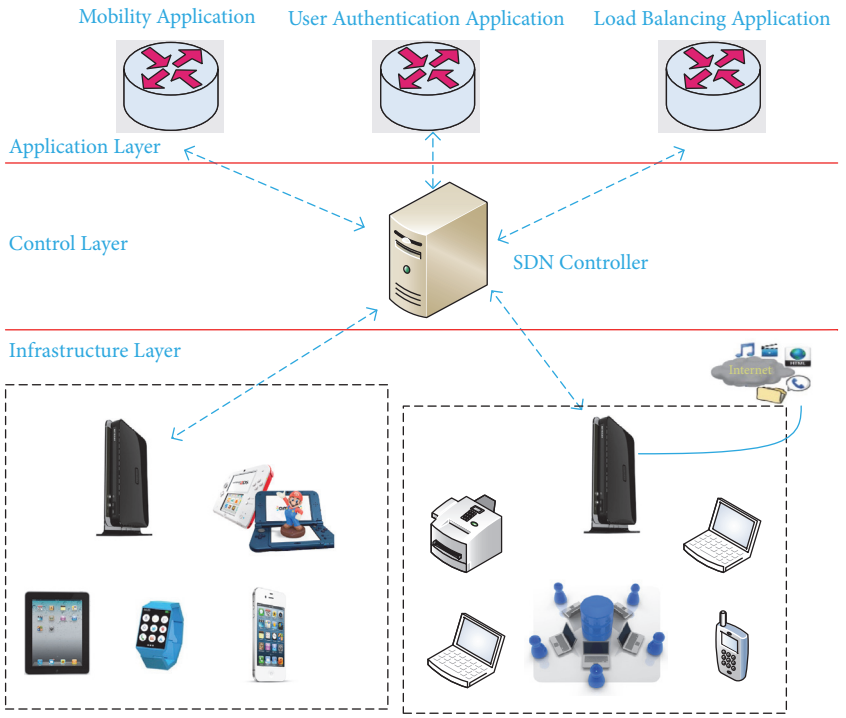


Figure 1: A software defined WiFi networking architecture.

VAP. Thereafter, this end-user can move freely within the network.

SDN provides a promising way to manage network in an elastic and cost-effective way. In recent years, the SDN has been developed beyond wired networks. We have witnessed various research efforts in implementing software defined wireless and cellular networks [12, 13]. After the separation of control and data plane customized configuration is no longer required for wireless APs. Due to this centralized nature, experiments can be run on the production network without generating traffic disturbance [14]. Network virtualization has also been examined to support wireless networks. Wireless network virtualization provides opportunities for several virtual networks to run on a shared wireless physical medium. It is promising to implement software defined wireless local area networks, while there are many challenges which need to be tackled before their widespread deployment, such as dynamically scheduling the wireless resources. With the increasing wide-range deployment and the diversification of wireless technologies, it has become a very challenging task to manage wireless networks. The software defined wireless networking poses many challenges and brings forth several research efforts to introduce the SDN features into wireless local area networks. Each access point conventionally makes decisions on its own modulation schemes, and power and channel settings based on local SNR evaluate or simply follow the default values.

The control plane of the WiFi infrastructure is much more complicated than the wired networks [12]. The first effort on instrumenting the control plane is the OpenRoads project [15], where a three-layer architecture was proposed. The flow layer implements the OpenFlow protocol to forward wireless traffic between routers. Medium specific parameters are managed using SNMP. The flow layer enables slicing traffic in order to allow an easy integration of new technologies and feasible experimentation on real networks. OpenRoads implements a similar approach like FlowVisor in wired networks [16]. The control layer centrally controls the network using the NOX controller [17] and the client can switch connections between cellular and WiFi networks to achieve seamless handovers. There are many advantages for a user to achieve enhanced coverage area and an increase in bandwidth capacity. Implementing the OpenRoads architecture requires decoupling mechanisms between service providers and network owners. This decoupling and virtualization over the laid infrastructure have far reaching effects in terms of economy and regulatory challenges faced by the industry [18]. Practical implementation of this architecture requires decoupling of service providers and network owners. OpenRadio fills this gap with the aim of providing programmability of the PHY and MAC layers by attempting to define a software abstraction layer that hides the hardware details from the upper layer programmers [19]. OpenRadio does not provide programmable PHY and MAC layers; nevertheless, it can cooperate with other projects such as WARP and CloudMAC. CloudMAC is a network architecture aimed at achieving a programmable MAC layer without resorting to software radios [20] with the introduction of the virtualized APs.

The deployment of SDWN for enterprise was studied in the Odin project [11], in which the light virtual access point (LVAP) approach was proposed, similar to LVAPs used in CloudMAC [20]. Odin and OpenRoad contribute a complete SDWN architecture. Nevertheless, there is still room for improving network delay and performance [21].

The channel-related processing may be a time critical job and the centralized processing away from the production network may significantly degrade delay sensitive applications such as VoIP or video streaming. It is not straightforward to apply the centralized control of SDN to wireless networks. Aeroflux steps up and tries to tackle the problem in a two-tier approach. AeroFlux was built based on the Odin framework [11]. This architecture divides the control plane in two layers. The lower layer, handled by nearsighted controllers (NSCs), is liable for situations that do not require global state data or those events that occur very frequently [22]. The services like load balancing and network monitoring, which are controlled by a central authority, are controlled by the global controller.

CloudMAC [23] is another software defined wireless network prototype in which APs are responsible for forwarding MAC frames. In this architecture, MAC frames are processed on the servers in data centers. CloudMAC WTPs require a WLAN driver and a small application for controlling infrastructure, which reduces software bugs and software complexity. Behop [24] is another SDWN architecture used for a wide set of management modules of channel, power, and association control in different environments. Utilizing the VAP abstraction to decouple WiFi logic from the physical infrastructure and control, the infrastructure is exposed to users. Behop also runs alongside production networks. Behop APs serve as OpenFlow switches and extend SDN functionalities to expose primitives for the channel, power, and association control.

In [25–27], authors proposed different SDWN architectures which utilize OpenWrt based embedded systems. In [25], Lee et al. developed an access point using Raspberry Pi. In [26] the instrumented SDWN platform can control channel assignment and interference management. In [27], Sundaresan et al. implemented and evaluated the performance of wireless home routers, in which the results showed how the characteristics of home wireless networks affect the performance of user traffic in real home environments. In [28, 29], the authors also utilized OpenWrt based systems and both architectures slice their network bandwidth in a software defined approach.

Our study includes various experiments to evaluate and compare the performance of our testbed with the existing WiFi infrastructures. The testbed is equipped with the latest OpenWrt firmware, packages, modifications in the specified modules, and drivers. Our testbed does not require any client-side modifications and our approach also removes the handoff delay with different channels in multi-BSS scenarios [30]. Previous studies considered different approaches towards the SDWN architectures with different scopes. Our testbed specifically focused on the handover performance of software defined WiFi networks with different parameters (such as number of clients, VAP, and packet size) in order to push the loading stress on single and multi-BSS scenarios [30–32].

3. Testbed

In order to enable programmability in the WiFi infrastructure, we construct our testbed without any client-side modifications. The network performance is evaluated with different workloads and types of traffic on the testbed. Distance and interference are the factors which have major impact on the throughput. The testbed includes updated versions of the Odin architecture with upgraded control functionalities, different applications running on the controller, the upgraded OpenWrt system, and the utility modules.

As shown in Figure 2, our testbed consists of three major parts. The first part is an SDN-based controller which centrally controls the whole network through different policy based applications. The second part includes a number of commodity access points, in which the NetGear routers (WNDR3700v4) serve as OpenFlow switches by instrumenting an OpenWrt based operating system. The OpenWrt release 15.05 is used for implementing the OpenWrt based image for the embedded Linux system. Different utilities are installed in order to make the devices function as OpenFlow switches. Major open source projects used in our testbed include the OpenVswitch version 2.3, the ath9k Linux driver, and the user-level click modular router. The NetGear WNDR 3700v4 model is equipped with the Atheros AR8327 chipset, 560 MHz CPU and 128 Mbit RAM. TP-Link TL-SG1024DT switches are used in order to provide the SSH utility and the Internet access to end-users. The third part includes 8 mini PCs which install the Intel Core i3 processor with 4 GB of RAM and 18 wireless USB adapters are used to serve as WiFi end-hosts. In order to send traffic from the iPerf clients to the iPerf server in a controlled manner, we utilized the clusterssh utility. Clusterssh provides the utility to issue the same command into several end-hosts in parallel. Otherwise, we have to log in each end-host with SSH and configure these hosts serially from a single input window over a SSH connection. Our experiments only test the 802.11 network at the 2.4 GHz range.

We also set up a multiple BSS network topology to study the handover performance when WiFi clients roam between APs as shown in Figure 3, in which there are two APs which are controlled by a SDN controller. Initially, a client is connected to AP-1. After some time this client moves its location towards AP-2. A laptop is used here for the handover experiments. In this study, dynamic IPs are assigned to the laptop by the TP-Link router. For displaying measurement results and maintaining network connectivity, we instrumented scripts to send ICMP messages periodically from one node to another.

4. Results

We conduct various measurement sessions and report the results of our SDWN testbed in this section. Our experiments were conducted with the following operational settings. Our testbed consists of multiple clients which are connected to the network, a master node which is responsible for the LVAP assignment and the installation of OpenFlow forwarding rules, and different applications which run on the master node. In order to send traffic at a time from different iPerf clients to the iPerf server, we utilize the clusterssh utility. Clusterssh provides the console to issue the same commands into several machines in a batch; otherwise, we have to log in each node with a standard SSH utility and issue the

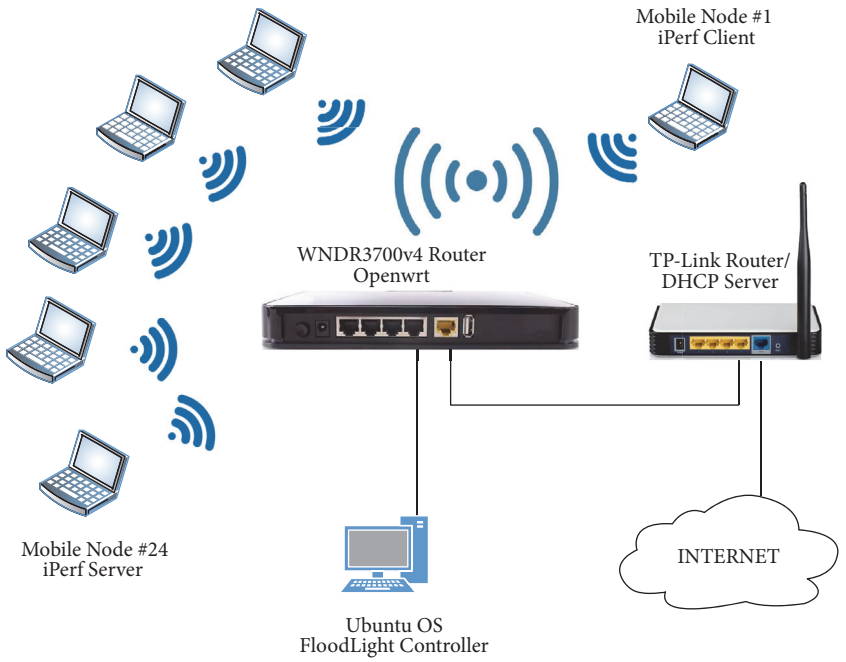


Figure 2: A single BSS network topology.

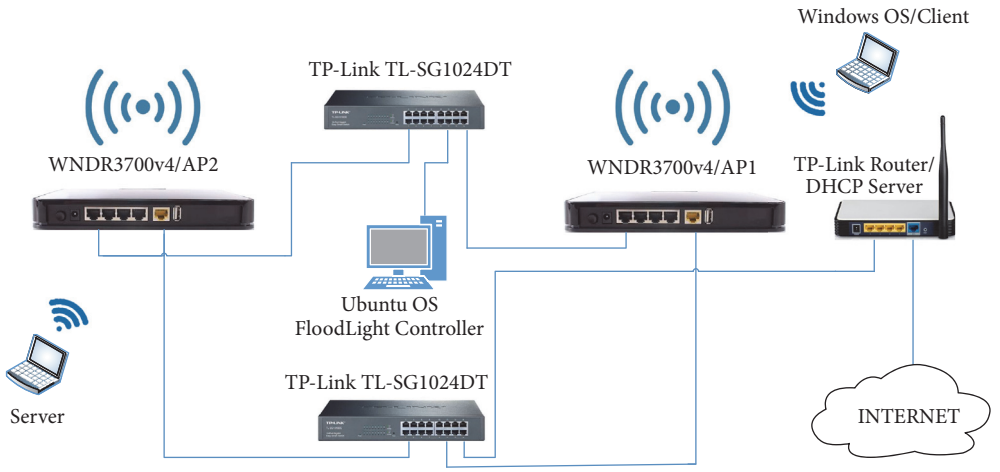


Figure 3: A multi-BSS network topology.

commands serially from a single input window over an SSH connection.

This testbed not only has instrumented the programmable WiFi nodes towards a software defined WiFi network, but also includes different techniques to improve the network performance. These experiments are designed to provide some insight into understanding the practicality of software defined systems in different aspects. In order to examine the design space of the software defined systems, we also instrumented a traditional WiFi network as the benchmark for the comparison purpose. Every access point has its own architecture or mechanisms; hence, its performance varies for association and reassociation with the clients. Without explicit statements, we use the same lab settings with different architectures in order to achieve a fair comparison. In this measurement study, we repeat our experiment sessions 10 times for each combination of the parameters, the reported values are the computed as the average of these 10 experiments, and MATLAB is used for plotting of result figures. We built this testbed with a simple NetGear WNDR 3700v4 switch. First, we measured the performance with the commodity hardware. Then, the OpenWrt based image was installed in the same NetGear switches for comparison with the Odin-V2 architecture. This Odin-V2 architecture is built based on OpenWrt. We aim to study whether the performance degradation was due to

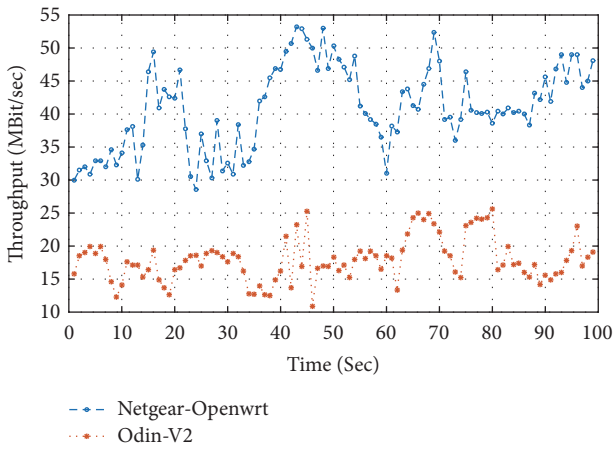


FIGURE 4: Network throughput with packet size $L = 1500$ bytes: NetGear-OpenWrt versus Odin-V2.

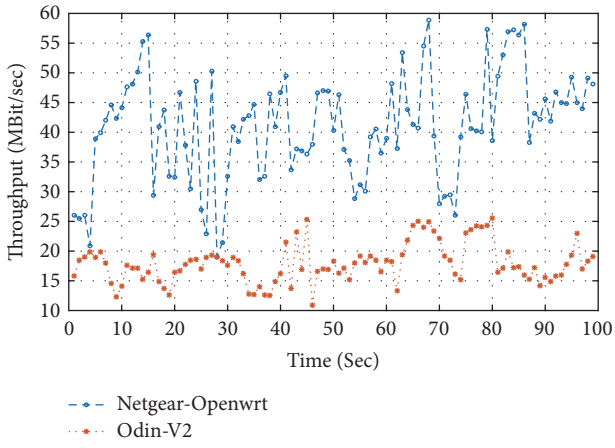


FIGURE 5: Network throughput with packet size $L = 1500$ bytes: generic NetGear versus Odin-V2.

OpenWrt or due to software implementation. Our evaluation starts with the throughput performance with different packet sizes and different topology settings.

Figure 2 shows the topology of our network testbed with a single BSS. In the first experiment, we set up a single access point with 2 clients, in which one is serving as the client and the other serves as the server. These two clients are associated with the same AP. One client is generating the packet streams and the other client receives the packets. The performance evaluation was conducted for 100 seconds with the packet size 1500 bytes in each session.

As shown in Figures 4 and 5, the throughput performance of Odin-V2 with the commodity NetGear WNDR 3700v4 switch and the NetGear WNDR 3700v4 with the installed OpenWrt firmware. The generic NetGear and the OpenWrt firmware were used to examine the delay performance benchmark to evaluate the software defined testbed based on OpenWrt. We aim to investigate that the performance degradation of software defined approach is either due to OpenWrt or some other reasons. The generated traffic is the overall network traffic because there are only two clients in

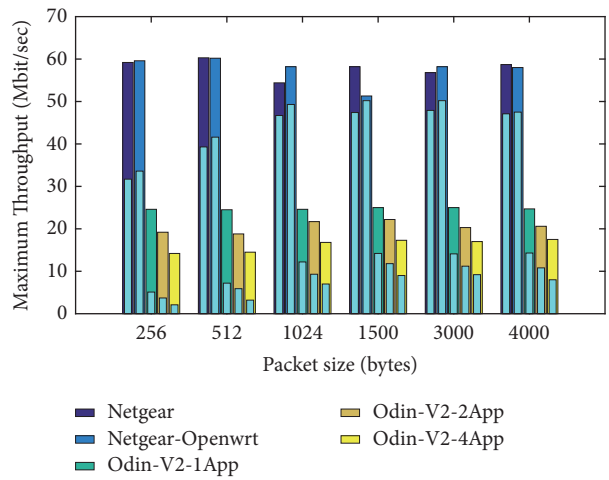


FIGURE 6: Maximum network throughput with 1 client: UDP versus TCP. Outer bar (blue bar) depicts the throughput for UDP flows and the inner bar shows the throughput of TCP flows.

this network. Afterwards, we increase the number of clients in the same topology. Among these clients, one client becomes the receiver or the iPerf server and the other remaining clients will generate traffic or behave as iPerf clients. All the traffic will be sent simultaneously towards the iPerf server by using the clustersh utility. In our results, outer bars depict the throughput for UDP flows and the inner bar shows the throughput of TCP flows.

Figure 6 illustrates the maximum network throughput of 1 client with UDP and TCP flows. This figure also contains the comparison of conventional network and Odin-V2 which includes 4 applications running on the controller. These applications create virtual access points. The applications running atop master include load balancing, authentication, and mobility. We also investigated the impact on network throughput by generating multiple slices on a single router. This kind of virtual access points can run on single or multiple access points within the same network depending on user's desire or need. The maximum throughout for the traditional network reaches around 60 Mbps; nevertheless, the performance of software defined approach is almost half of the conventional network. This performance drop is due to controlled traffic in SDN because multiple LVAPs are created. Other reasons may be the software implementation of switching in AP. The software defined approach running with multiple applications has less throughput as compared to single application running on controller. For each individual client, an individual virtual AP is created to deploy specific set of packet process rules. Figures 7 and 8 illustrate the performance of the 6 and 12 clients for UDP and TCP flows with respect to packet sizes. In these scenarios, we increased the traffic generating clients from 6 to 12. The measurement results show that Odin-V2 network is approaching 15 Mbps for smaller packet sizes and the average network throughput is more than half for cases with smaller packet sizes. In case of larger packet sizes, the throughput difference between UDP and TCP flows drops down to half. The relationship between

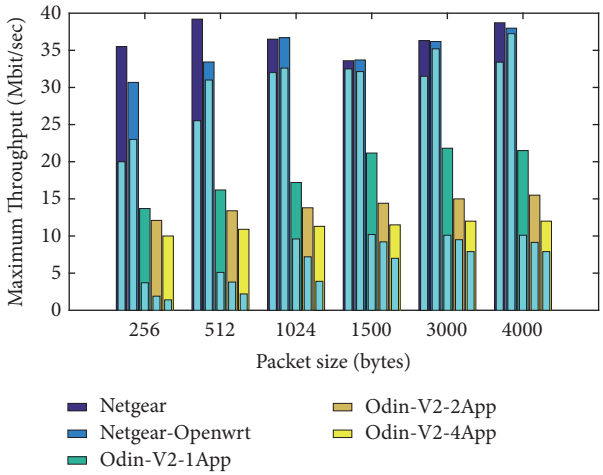


Figure 7: Maximum network throughput with 6 clients: UDP versus TCP. Outer bar (blue bar) depicts the throughput for UDP flows and the inner bar shows the throughput of TCP flows.

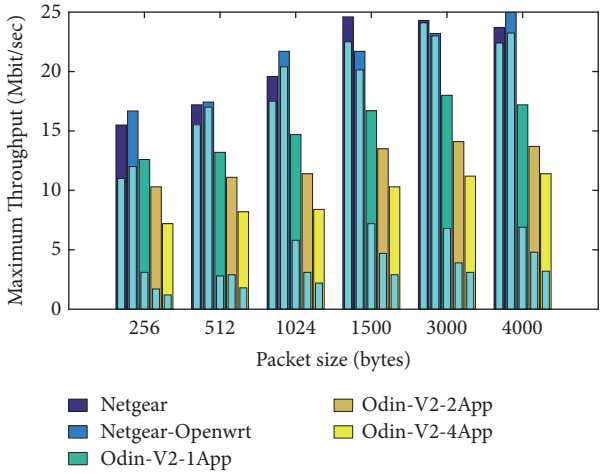


Figure 8: Maximum network throughput with 12 clients: UDP versus TCP. Outer bar (blue bar) depicts the throughput for UDP flows and the inner bar shows the throughput of TCP flows.

throughputs versus packet size is proportional. However, when the packet size is too large, it becomes difficult to deliver the packets to the receiver side. In case of noisy channel conditions and large packet sizes, the throughput reduces more than previous conditions due to increasing retransmission. Usually one AP can provide services to about 25 clients at a time but it also depends on the AP architecture. So in order to test the network performance under higher stress conditions, we utilize around 24 clients. Figure 9 also illustrates the performance evaluation of 24 clients with both architectures. We observed that Odin-V2 with different applications is still working. The TCP throughput is around 5 Mbps and the UDP throughput is less than 10 Mbps with different packet sizes but even with this throughput many multimedia applications can work well within this range.

Generally speaking, all the devices in a WiFi network do not remain in the active mode simultaneously. These devices

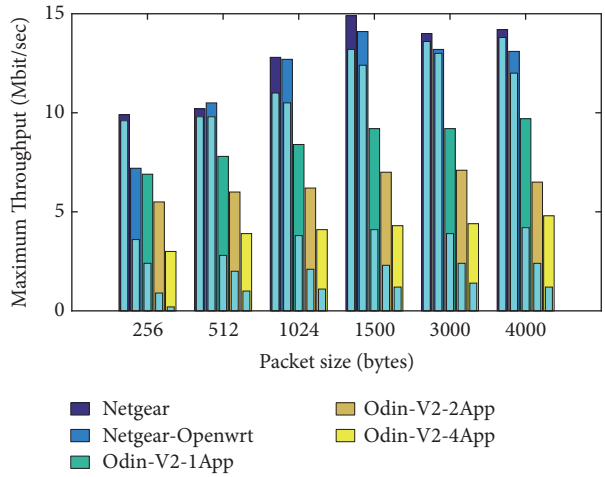


Figure 9: Maximum network throughput with 24 clients: UDP versus TCP. Outer bar (blue bar) depicts the throughput for UDP flows and the inner bar shows the throughput of TCP flows.

switch as active and passive clients from time to time within the network. In Figures 10 and 11, we aim to study the impact of passive clients on the network throughput. Hence, we increased the number of clients from 6 to 24 clients by having half of the clients in the passive mode. These results show the falling trend of throughput of about 6 to 10 Mbps in case of passive clients. Figure 12 also illustrates the same performance degradation trend for different types of flows with constant 1500 bytes packet size. The primary drop in performance of the conventional WiFi network can be observed during the increment in the number of clients from 1 to 6.

Figure 13 shows the throughput comparison of all active clients with different combination of active and passive clients. The inner bar depicts the performance of all active clients and the outer bar shows the performance of combination of active and passive clients. These experiments demonstrate the impact of the increasing number of passive clients on the throughput of active clients. These results also show that the conventional network has more impact on the performance degradation as compared to the software defined approach in case of a small number of clients.

Figure 14 shows the impact on the throughput and the reassociation delay during handover. Generally handover occurs due to signal strength. When a client receives better signal strength from an AP, as compared to previously associated AP, it reassociates itself towards another AP. This handover mechanism is not specified in the 802.11 protocols. It depends on the client architecture how it behaves under this specific situation. There are two major factors behind the variation of handover delay: one is the delay occurrence due to AP and the other is due to initialization of client. In our study, a client initiates this procedure. In order to minimize the effect of different architectures, we have utilized the same hardware for all cases. In this experiment, the client moves from one AP towards another after approximately 25 seconds. The results show that the software defined approach has negligible fluctuation during handover. One reason behind

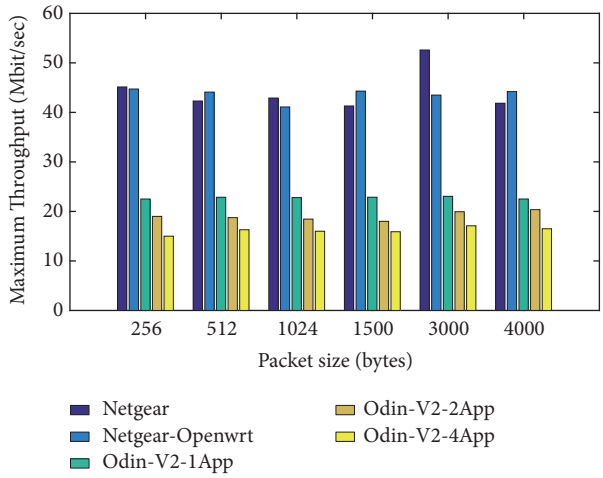


Figure 10: Maximum network throughput with 3 active and 3 passive clients.

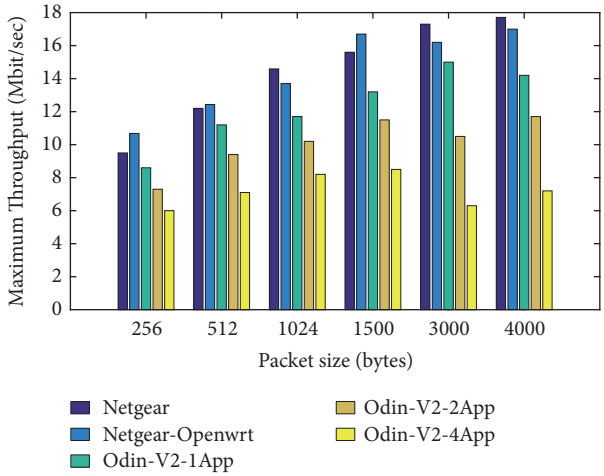


Figure 11: Maximum network throughput with 12 active and 12 passive clients.

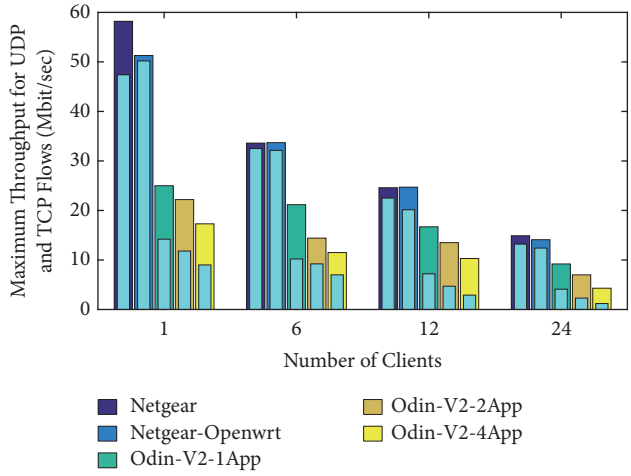


Figure 12: Network throughput with packet size $L = 1500$ bytes. Outer bar (blue bar) depicts the throughput for UDP flows and the inner bar shows the throughput of TCP flows.

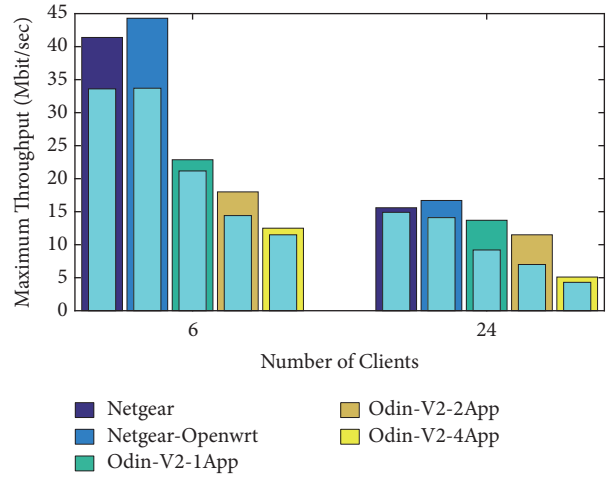


Figure 13: Network throughput with different number of active and passive clients. The inner bar depicts the performance of all active clients and the outer bar shows the performance of combination of active and passive clients.

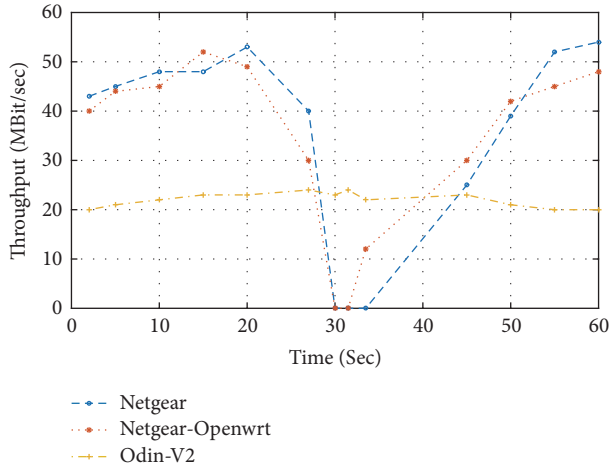


Figure 14: Network throughput during the handover.

this normal transition state is that there is no exchanging of layer-2 and layer-3 messages. The other reason is due to the centralized control, whereas the conventional network takes around 2 to 3 seconds for reassociation and takes more time to regain the prehandover throughput. Although Odin-V2 provides less throughput as compared to the conventional network, this handover delay provides a practical solution for many streaming applications, where minor delay can affect the performance of applications, such as VoIP and online gaming. Figure 15 shows the round-trip time, where the purpose of this experiment is to find the end-to-end latency, because it has a negative effect on throughput. For small packet sizes, the end-to-end latency is almost the same but for large packet sizes Odin-V2 show 4 times more latency as compared to conventional network. Figure 16 shows the handover delay; in this case the two access points have different channels. As mentioned earlier, Odin-V2 does not exchange layer-2 and layer-3 messages between clients

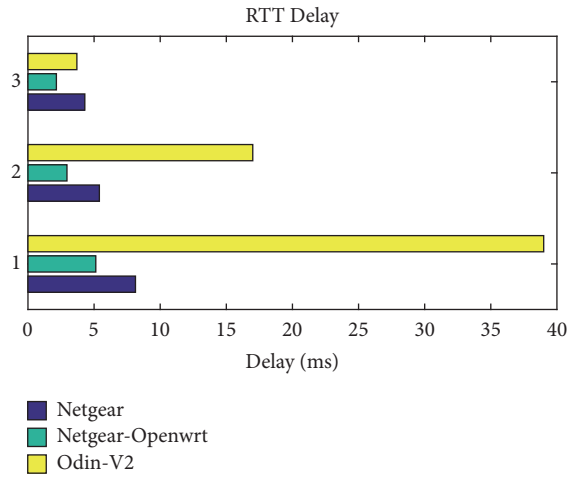


Figure 15: The round-trip time performance comparison.

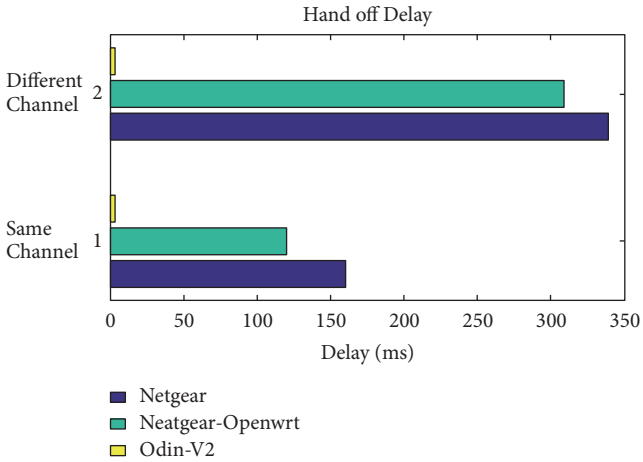


Figure 16: Handover delay on the same operation channel and the different operation channels.

and APs, and the handover delay is practically negligible compared with the conventional WiFi network. However, the small delay as shown in Figure 16 is due to the communication between the AP and the controller, which is responsible for association/dissociation of a client from one AP to another.

With lacking of coordination between APs, a traditional WiFi network can experience unbalance traffic load. Figures 17 and 18 illustrate two scenarios, before and after load balancing with different number of access points. These figures depict a hypothetical scenario for balancing network load among different access points, showing how SDN helps to balance its load in order to improve the network throughput among different clients. This hypothetical scenario is tabulated in Table 1, showing the scenario of three access points, and Table 2 shows the scenario of 4 access points.

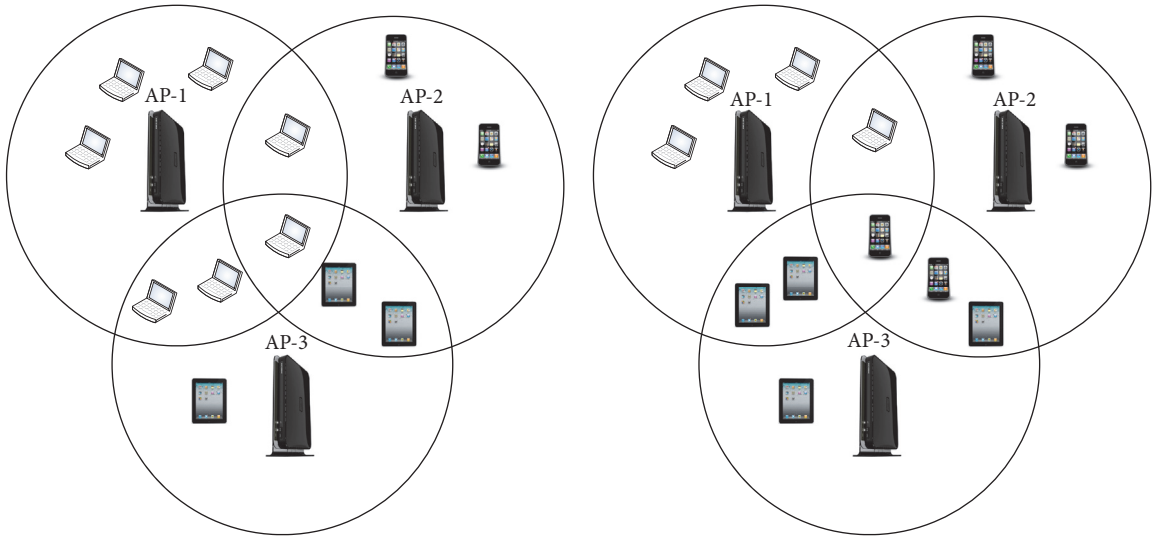
In summary, Figure 19 shows the tradeoff between programmability and performance. Odin-V2 shows lower throughput as compared to the traditional WiFi network devices but does also show the flexibility of a programmable device. The throughput difference between the NetGear-OpenWrt implementation and the SDN-based wireless network is due to the OpenVswitch and the click modular router because the click modular router is running at the user space of Linux. The degradation on the throughput performance by SDN is tolerable because this VAP implementation provides a smooth handover, which indicates the user connection migration from one AP to another without any noticeable delay.

5. Conclusion

In this paper, we conducted a measurement study of an SDWN testbed with different packet sizes, virtual access points, and number of clients to investigate how different SDN-based WiFi networks behave with different traffic loads with typical network settings. We investigated the handover mechanism of SDWN and traditional WiFi networks by the deployment of multiple access points. Our practical deployment experiences depicted the behaviors of our SDWN infrastructure with several applications running on the controller. We also observed the performance implication of deploying different applications on the controller in SDWN. The average and maximum throughput performance of TCP and UDP flows was evaluated for different network testing scenarios. Our testbed platform, Odin-V2, is based on OpenWrt; hence, the performance of Odin-V2 was therefore gauged using OpenWrt as the benchmark.

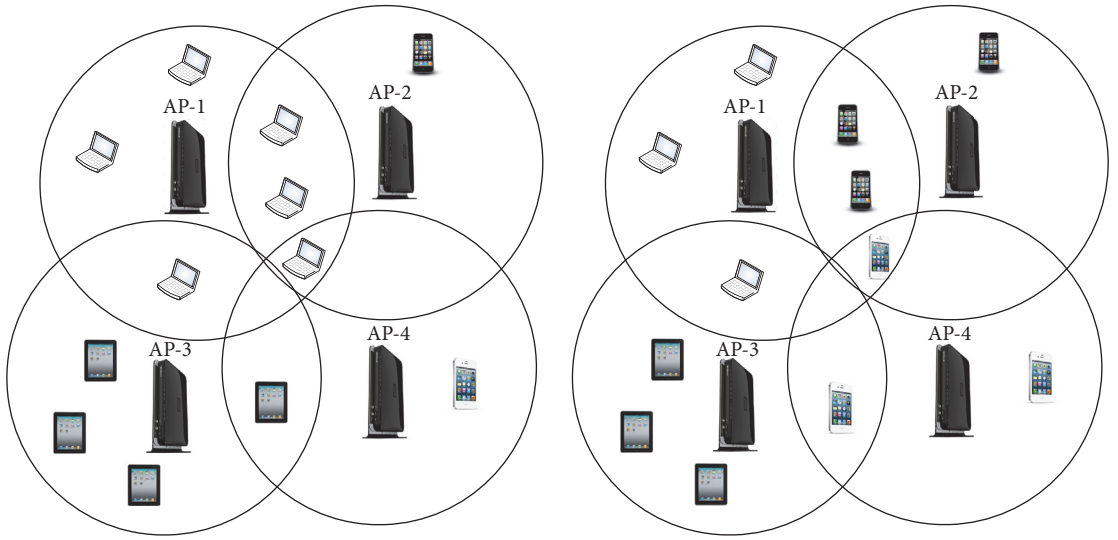
The logically centralized nature of the SDWN provides many benefits for management at the cost of performance degradation. In particular, the performance with a large number of clients is still an open issue. SDWN may be a promising approach to solve many issues including handover, load balancing, and managing complexity. However, before large-scale deployment of this software defined approach, several issues including throughput, delay, resource discovery, and security need to be addressed. Latency in SDWN is still an open research challenge for many applications which is also demonstrated by our study. The SDWN architecture provides a rich set of control features, while traditional WiFi networks are still advantageous in better network performance. This comparison study of different SDWN implementations shows that no individual architecture can fulfill the demand of users, so network administrator should devise application specific architectures for network optimization. We conjecture that the performance degradation of this software defined approach is due to the click modular router used in our testbed. Such a software-based implementation in the user space of Linux is a major issue; system overhead may also arise due to generation of different LVAPs created for each user.

In order to gain full advantages of the SDWN architectures, specific routers can be designated for specific applications. This would increase the share of deterministic component of the overall load compared to stochastic nature for the rest of the load. As deterministic models are easier to predict, it makes them more reliable compared to their stochastic counterparts. The programmable nature of the proposed SDWN architecture also enables us to counter for the inherent drawbacks of a deterministic model. This can be achieved by scheduling certain applications to start on certain



(a) Before (b) After

Figure 17: Load balancing with 3 access points.



(a) Before (b) After

Figure 18: Load balancing with 4 access points.

Table 1: Load balancing case I.

Device	Without load balancing			With load balancing		
	AP1	AP2	AP3	AP1	AP2	AP3
Generic NetGear	8.32 Mbps	19.43 Mbps	29.11 Mbps	8.31 Mbps	19.42 Mbps	29.13 Mbps
NetGear-OpenWrt	7.31 Mbps	8.31 Mbps	25.6 Mbps	7.31 Mbps	8.32 Mbps	25.61 Mbps
Odin-V2	3.52 Mbps	8.33 Mbps	12.52 Mbps	6.25 Mbps	6.25 Mbps	6.25 Mbps

Table 2: Load balancing case II.

Device	Without load balancing				With load balancing			
	AP1	AP2	AP3	AP4	AP1	AP2	AP3	AP4
Generic NetGear	9.7 Mbps	14.55 Mbps	58.2 Mbps	58.2 Mbps	9.7 Mbps	14.55 Mbps	58.2 Mbps	58.2 Mbps
NetGear-OpenWrt	8.5 Mbps	12.82 Mbps	51.3 Mbps	51.3 Mbps	8.5 Mbps	12.82 Mbps	51.3 Mbps	51.3 Mbps
Odin-V2	4.16 Mbps	6.25 Mbps	25 Mbps	25 Mbps	8.33 Mbps	8.33 Mbps	8.33 Mbps	8.33 Mbps

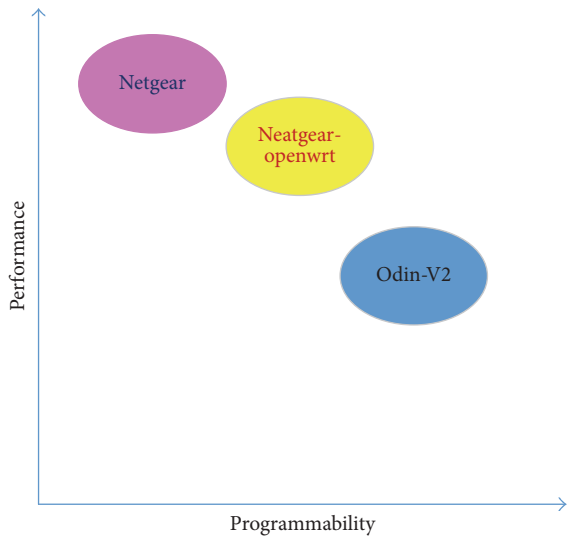


Figure 19: Performance versus programmability.

routers based on variety of parameters including users and time. The network load can be further balanced and thus the network performance may be improved by introducing hybrid traffic balancing of passive and active modes. The throughput performance of the proposed network model can also be further improved using better network management policies. We plan to design and implement intelligent load balancing schemes on Zynq-based programmable WiFi systems in a software/hardware codesign approach [33].

Conflicts of Interest

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

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (no. 61370231).

References

[1] "Cisco Visual Networking Index: Forecast and Methodology," 2017, https://www.cisco.com.

[2] N. Feamster, J. Rexford, and E. Zegura, "An intellectual history of programmable networks," *Queue*, vol. 11, no. 12, Article ID 2560327, 2013.

[3] T. Zahid, F. Y. Dar, X. Hei, and W. Cheng, "An empirical study of the design space of smart home routers," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 9677, pp. 109–120, 2016.

[4] A. Lara, A. Kolasani, and B. Ramamurthy, "Network innovation using open flow: a survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 493–512, 2014.

[5] Y. Gao, L. Dai, and X. Hei, "Throughput Optimization of Multi-BSS IEEE 802.11 Networks with Universal Frequency Reuse," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3399–3414, 2017.

[6] Y. Zhang, "GroRec: a group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[7] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. PP, no. 99, 2015.

[8] M. U. Aslam, A. Derhab, K. Saleem et al., "A Survey of Authentication Schemes in Telecare Medicine Information Systems," *Journal of Medical Systems*, vol. 41, no. 1, article no. 14, 2017.

[9] C. Xu, Q. Chen, H. Hu, J. Xu, and X. Hei, "Authenticating Aggregate Queries over Set-Valued Data with Confidentiality," *IEEE Transactions on Knowledge and Data Engineering*, 2017.

[10] S. Sezer, S. Scott-Hayward, P. Chouhan et al., "Are we ready for SDN? Implementation challenges for software-defined networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 36–43, 2013.

[11] L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, and T. Vazao, "Towards programmable enterprise WLANS with Odin," in *Proceedings of the 1st ACM International Workshop on Hot Topics in Software Defined Networks (HotSDN '12)*, pp. 115–120, ACM, Helsinki, Finland, August 2012.

[12] C. J. Bernardos, A. de la Oliva, P. Serrano et al., "An architecture for software defined wireless networking," *IEEE Wireless Communications Magazine*, vol. 21, no. 3, pp. 56–61, 2014.

[13] Y. Zhang, M. Chen, N. Guizani, D. Wu, and V. C. Leung, "SOVCAN: Safety-Oriented Vehicular Controller Area Network," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 94–99, 2017.

[14] Y. Yiakoumis, K.-K. Yap, S. Katti, G. Parulkar, and N. McKeown, "Slicing home networks," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Home Networks (HomeNets '11)*, pp. 1–6, August 2011.

[15] K.-K. Yap, M. Kobayashi, R. Sherwood et al., "OpenRoads: Empowering research in mobile networks," *SIGCOMM Computer and Communications*, vol. 40, no. 1, pp. 125-126, 2010.

[16] R. Sherwood, M. Chan, and A. Covington, "Carving research slices out of your production networks with openflow," *Computer Communication Review*, vol. 40, no. 1, pp. 129-130, 2010.

[17] N. Gude, T. Koponen, and J. Pettit, "NOX: towards an operating system for networks," *Computer Communication Review*, vol. 38, no. 3, pp. 105–110, 2008.

[18] C. E. Rothenberg, M. R. Nascimento, M. R. Salvador, C. N. A. Corrêa, S. Cunha De Lucena, and R. Raszuk, "Revisiting routing control platforms with the eyes and muscles of software-defined networking," in *Proceedings of the 1st ACM International Workshop on Hot Topics in Software Defined Networks, HotSDN 2012*, pp. 13–18, Finland, August 2012.

[19] M. Bansal, J. Mehlman, S. Katti, and P. Levis, "OpenRadio: a programmable wireless dataplane," in *Proceedings of the 1st ACM International Workshop on Hot Topics in Software Defined Networks (HotSDN '12)*, pp. 109–114, Helsinki, Finland, August 2012.

[20] P. Dely, J. Vestin, A. Kassler, N. Bayer, H. Einsiedler, and C. Peylo, "CloudMAC - An OpenFlow based architecture for 802.11 MAC layer processing in the cloud," in *Proceedings of the 2012 IEEE Globecom Workshops, GC Wkshps 2012*, pp. 186–191, USA, December 2012.

[21] B. Heller, R. Sherwood, and N. McKeown, "The controller placement problem," in *Proceedings of the 1st ACM International*

Workshop on Hot Topics in Software Defined Networks, HotSDN 2012, pp. 7–12, Finland, August 2012.

[22] J. Schulz-Zander, C. Mayer, B. Ciobotaru, S. Schmid, and A. Feldmann, "OpenSDWN: programmatic control over home and enterprise WiFi," in *Proceedings of the the 1st ACM SIGCOMM Symposium on Software Defined Networking Research (SOSR '15)*, pp. 1–12, Santa Clara, Calif, USA, June 2015.

[23] J. Vestin, P. Dely, A. Kassler, N. Bayer, H. Einsiedler, and C. Peylo, "CloudMAC," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 4, p. 42, 2013.

[24] Y. Yiakoumis, M. Bansal, A. Covington, J. Van Reijendam, S. Katti, and N. McKeown, "BeHop: A testbed for dense WiFi networks," in *Proceedings of the 9th ACM MobiCom Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization, WiNTECH 2014*, pp. 1–8, USA, September 2014.

[25] W. J. Lee, J. W. Shin, H. Y. Lee, and M. Y. Chung, "Testbed implementation for routing WLAN traffic in software defined wireless mesh network," in *Proceedings of the 8th International Conference on Ubiquitous and Future Networks, ICUFN 2016*, pp. 1052–1055, Austria, July 2016.

[26] R. Riggio, M. K. Marina, and T. Rasheed, "Interference management in software-defined mobile networks," in *Proceedings of the 14th IFIP/IEEE International Symposium on Integrated Network Management, IM 2015*, pp. 626–632, Canada, May 2015.

[27] S. Sundaresan, N. Feamster, and R. Teixeira, "Measuring the performance of user traffic in home wireless networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 8995, pp. 305–317, 2015.

[28] K. L. Huang, C. L. Liu, C. H. Gan, M. L. Wang, and C. T. Huang, "SDN-based wireless bandwidth slicing," in *Proceedings of the International Conference on Software Intelligence Technologies and Applications & International Conference on Frontiers of Internet of Things*, pp. 77–81, Hsinchu, Taiwan, 2014.

[29] H. H. Gharakheili, L. Exton, and V. Sivaraman, "Managing home routers from the cloud using Software Defined Networking," in *Proceedings of the 13th IEEE Annual Consumer Communications and Networking Conference, CCNC 2016*, pp. 262-263, USA, January 2016.

[30] T. Zahid, X. Hei, and W. Cheng, "Understanding performance bottlenecks of a multi-BSS software defined WiFi network testbed," in *Proceedings of the 1st IEEE International Conference on Computer Communication and the Internet, ICCCI 2016*, pp. 153–156, China, October 2016.

[31] T. Zahid, F. Y. Dar, X. Hei, and W. Cheng, "A measurement study of a single-BSS software defined WiFi testbed," in *Proceedings of the 1st IEEE International Conference on Computer Communication and the Internet, ICCCI 2016*, pp. 144–147, China, October 2016.

[32] T. Zahid, X. Hei, and W. Cheng, "Understanding the Design Space of a Software Defined WiFi Network Testbed," in *Proceedings of the 14th International Conference on Frontiers of Information Technology, FIT 2016*, pp. 170–175, Pakistan, December 2016.

[33] J. Kang, X. Hei, and J. Song, "A Comparative Study of Zynq-Based OpenFlow Switches in a Software/Hardware Co-design," in *International Workshop on Network Optimization and Performance Evaluation (NOPE)*, vol. 10658 of *Lecture Notes in Computer Science*, pp. 369–378, Springer International Publishing, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 8263704, 9 pages
https://doi.org/10.1155/2018/8263704

Research Article

A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework

Yibo Wang,[1] Mingming Wang,[1] and Wei Xu [1,2]

[1]*School of Information, Renmin University of China, Beijing 100872, China*
[2]*Smart City Research Center, Renmin University of China, Beijing 100872, China*

Correspondence should be addressed to Wei Xu; weixu@ruc.edu.cn

Received 2 December 2017; Accepted 3 January 2018; Published 22 March 2018

Academic Editor: Yin Zhang

Movie recommendation in mobile environment is critically important for mobile users. It carries out comprehensive aggregation of user's preferences, reviews, and emotions to help them find suitable movies conveniently. However, it requires both accuracy and timeliness. In this paper, a movie recommendation framework based on a hybrid recommendation model and sentiment analysis on Spark platform is proposed to improve the accuracy and timeliness of mobile movie recommender system. In the proposed approach, we first use a hybrid recommendation method to generate a preliminary recommendation list. Then sentiment analysis is employed to optimize the list. Finally, the hybrid recommender system with sentiment analysis is implemented on Spark platform. The hybrid recommendation model with sentiment analysis outperforms the traditional models in terms of various evaluation criteria. Our proposed method makes it convenient and fast for users to obtain useful movie suggestions.

1. Introduction

The popularity of mobile devices makes people's daily lives more dependent on mobile services. People get business information, product information, promotion information, and recommendation information from mobile devices. An important application of mobile services is movie recommendation. A movie recommender system has proven to be a powerful tool on providing useful movie suggestions for users. The suggestions are provided to support the users in their effort to cope with the information overload and help them find appropriate movies fast and conveniently. Different from the demand on the personal computers (PCs), mobile services place more emphasis on timeliness, which requires fast processing and calculation from service providers. Therefore, movie recommendation in mobile services needs to be promoted in both the recommendation accuracy and the timeliness.

Movie recommendation is a comprehensive and complicated task which involves various tastes of users, various genres of movies, and so forth. Therefore, lots of techniques for recommendation have been proposed to solve the problems. For example, content-based recommender system, collaborative filtering recommender system, and hybrid recommender system. Each technique has its own advantage in solving specific problems. Considering the usage of online information and user-generated content, collaborative filtering is supposed to be the most popular and widely deployed technique in recommender system. Collaborative filtering method recommends items by measuring the similarity between users. The similarity between users' preference can be measured by correlation calculation. In this way, users who have similar interest in movies are sorted in the same group, and then movies are recommended by their reviews and ratings of movies that they have seen. However, the correlation and similarity are difficult to calculate due to the sparsity of user's basic data, such as users' rating on movies that they have watched and their browsing history. Actually, the reviews of users on movies usually contain more information such as users' preference. Moreover, the ignorance of sentiment which users have is also a big problem in movie recommendation. At present, people are increasingly willing to post their own reviews online. In their reviews, users can express their preferences and feelings about movies.

And the feelings contained in these reviews also affect the choice of other users. Users will see the reviews, analyze their personal experience, choose their useful reviews, remove some misleading or even harmful reviews, and ultimately make their own judgments and decisions. Therefore, the sentiment in reviews is a very important aspect in evaluating a movie. Generally speaking, users are more inclined to choose the movies that the majority of people prefer and abandon the movies that the majority of people dislike. The decisions are made according to other people's experience to achieve users' own comfort experience.

With the increase of the amount of data, how to provide users with high-quality recommendations quickly among the massive information has become a serious problem. The arrival of the mobile services makes the response speed an important indicator of the user experience. The text mining and sentiment analysis techniques used to deal with user reviews aggravate the difficulty of the recommender system in the traditional environment. A new generation of recommender system needs to address how to make high-quality recommendations quickly in massive amounts of data and how to make the system highly scalable. Big data technology is one of the powerful tools to solve these problems. Some recommender systems based on Hadoop can alleviate the calculation pressure caused by the increase in the amount of data. However, in the circumstance of complex process or large number of iterations, Hadoop is not an appropriate tool because of enormous I/O access. Extremely long processing time is a critical flaw for Hadoop under the requirement of high timeliness. Fortunately, the emergence of Spark meets these needs. Different from disk-based storage of Hadoop, Spark is more inclined to save the intermediate results in memory in the calculation process, and the iterative calculation process has also been optimized. So Spark's processing efficiency is better than Hadoop in recommender systems.

In this paper, a sentiment-enhanced hybrid collaborative filtering and content-based recommendation method is proposed to recommend appropriate movies to users on Spark platform. Sentiment analysis is more reliable than simple rating, due to the fact that it contains more emotional information, which proves to be powerful in arts items such as movies. Moreover, the high efficiency of Spark makes it possible to improve the timeliness of mobile services.

The remaining sections of this paper are organized as follows: Section 2 summarizes the existing research work. The sentiment-enhanced recommendation framework is proposed in Section 3. The empirical analysis and experimental results are shown in Section 4. Finally, conclusions and future work are given in Section 5.

2. Related Work

2.1. Movie Recommender Systems. A recommender system is a program that predicts users' preferences and recommends appropriate products or services to a specific user based on users' information and products or services information. The research on recommender systems is started by GroupLens research team from the University of Minnesota. Their research object is a movie recommender system called MovieLens. Early research is mainly focused on the content of the recommender system which analyzed the characteristics of the object itself to complete the recommendation task [1]. However, this recommendation method can only be confined to content analysis, which makes researchers and practitioners invest great efforts in designing new recommender systems. Researchers have proposed recommender systems based on collaborative filtering, association rules [2], utility, knowledge, social network [3], multiobjective programming [4], clustering [5], and other theories and techniques.

Researchers have also studied recommendations on mobile devices. Most of the research on mobile recommendation focuses on location-based services. For example, Zheng et al. utilized GPS trajectory data to solve mobile recommendation problems [6]. They proposed a user-centered collaborative location and activity filtering method based on user-location-activity relations and collaborative filtering recommendation method. On the basis of this study, Zheng et al. came up with an algorithm using ranking-based collective tensor and matrix factorization (MF) to recommend activities to users [7]. Moreover, Park et al. recommended users with restaurants using Bayesian networks based on location and some other information [8].

2.2. Content-Based Recommendation. Content-based movie recommendation methods have been widely explored in the past few years. Basu et al. proposed a content-based movie recommender system using ratings of the movies as the social information [1]. The experiments proved that their methods were more flexible and accurate. What is more, Ono et al. employed Bayesian networks to construct users' movie preference models based on their context [9]. Obviously, a variety of methods were used to excavate features of users and movies to recommend appropriate movies. In addition to use new technologies to explore features, new perspectives are also explored to build accurate profiles of users and movies. For example, Szomszor et al. introduced semantic web to analyze folksonomy hidden in the movies to help users discover appropriate movies [10]. De Pessemier et al. used social network to analyze the individual context features on users' purchasing behavior [11]. However, the design of effective profiles is always the bottleneck of content-based recommender systems. Both researchers and practitioners have made great efforts in designing a new recommendation method to avoid the shortcoming of content-based recommender systems.

2.3. Collaborative Filtering-Based Recommendation. Collaborative filtering is used to make up for the shortcomings of content-based algorithm [12]. The collaborative filtering algorithm was divided into parts for deep analysis in movie recommendation by Herlocker et al. [12]. In the process of recommendation, Koren found that users' preference changed over time, so he came up with a recommendation method using temporal dynamics to solve the problem [13]. What is more, Hofmann implemented Gaussian probabilistic latent semantic analysis in the collaborative filtering method on movie recommendation research [14]. Researchers invested

great efforts by adding new technologies to improve the performance of collaborative filtering methods on movie recommendation and they achieved good results.

Collaborative filtering is a prevalent tool used in recommender systems [15]. Marlin came up with a collaborative filtering method based on ratings [16]. Salakhutdinov and Mnih proposed a collaborative filtering method, Probabilistic Matrix Factorization, which can handle large scale of dataset [17]. At the same time, Salakhutdinov et al. employed Restricted Boltzmann Machines to improve the performance of collaborative filtering [18]. The experiment results showed that Restricted Boltzmann Machines outperformed singular value decomposition (SVD) on Netflix dataset. Moreover, Koren combined improved latent factor models and neighborhood models on Netflix dataset [19]. The latent factor model used is SVD while the neighborhood model is optimized on loss function. What is more, researchers also introduced other data mining methods to optimize the recommender systems. For example, Rendle proposed Factorization Machines (FM) which combine support vector machines (SVM) with factorization models [20]. Zhen et al. used the regularized MF used in Probabilistic Matrix Factorization (PMF) with tagging information of movies [21].

However, collaborative filtering method introduced new drawbacks in making up for some of the shortcomings of the content-based method. For example, the scalability of collaborative filtering is poor. When users produce new behavior, it is difficult for collaborative filtering to respond immediately. Therefore, both researchers and practitioners are inclined to hybridize collaborative filtering method and content-based method to solve the problem [22, 23]. For example, Debnath et al. presented a collaborative filtering and content-based movie recommender system [24]. In the content-based part of the hybrid system, the importance of the feature is expressed in a weighted manner. Nazim Uddin et al. proposed a diverse-item selection algorithm for optimizing the output of collaborative filtering method to improve the performance of hybrid recommender system [25]. Gunawardana and Meek introduced unified Boltzmann machines to hybrid collaborative filtering method and content-based method by encoding their information [26]. On the basis of integration of content-based method and collaborative filtering method, Soni et al. joined the analysis of review based text mining algorithm, making the recommendation more accurate [27]. Moreover, Ling et al. employed a rating model with a topic model based on reviews to make accurate predictions [28]. As can be seen from the above studies, the hybrid recommender system can not only improve the efficiency, but also improve the scalability of movie recommendation. Therefore, a hybrid recommendation model is an appropriate method of movie recommendation.

2.4. Sentiment Analysis. Sentiment analysis is the process of analyzing, processing, summarizing, and reasoning the emotional text [29]. Sentiment analysis began in 2002 by Pang et al.'s research [30] and has been greatly developed in the online commentary about the emotional polarity analysis. At present, the accuracy of emotional polarity analysis based on online commentary text is gradually increasing, but one of the problems existing in emotional analysis is the lack of in-depth analysis and application of the influence of sentiment analysis.

Pang et al. used supervised learning method in machine learning to classify emotional polarity of the movie commentary text into positive one and negative one, by using the part of speech (POS) N-gram grammar (n-gram) and maximum entropy (ME) [30]. Turney implemented the unsupervised learning of machine learning to study the polarity of the text emotion [31]. He first used tags to extract the word pair from reviews and then used Pointwise Mutual Information and Information Retrieval (PMI-IR) method to calculate the similarity between the words in the text and words in the corpus to determine the emotional polarity of the text. The commentary data come from the online comment site http://Epinions.com. The method obtained an accuracy of 65.83% in the movie reviews dataset.

The polarity of reviews of the movies and other goods or services can be divided into positive, negative, and neutral. In general, the researchers believe that positive information has a positive effect while negative information has a negative effect [32]. Based on this conclusion, some studies introduced sentiment analysis into the user's reviews and obtained the polarity of the reviews. Then movies with most positive information were recommended to users [33]. Sun et al. came up with a sentiment-aware social media recommender system [34]. Diao et al. analyzed sentiment of reviews in collaborative filtering by applying a topic model [35].

2.5. Big Data Analytics for Recommendation. The scalability problem of recommender system also makes it harder for researchers and practitioners to provide users with convenient and efficient services. Many efforts have been taken to solve the problem [36–38]. Parallel computing is one of the most prevalent solutions. Zhou et al. built a parallel Matlab platform to implement a movie recommender system with collaborative filtering method [39]. In parallel computing, the operation efficiency of recommendation algorithms is higher than that of single machine operation. The introduction of the distributed computing framework makes the efficiency of the recommender system improve qualitatively. For example, Hadoop could help the collaborative filtering method achieve linear speedup [40, 41]. And larger datasets could get a better speedup than smaller ones [42]. Although Hadoop alleviates the scalability of recommendation algorithms to some extent, the support of MapReduce for collaborative filtering algorithms is not perfect. The reason is that collaborative filtering requires constant reading and writing of data in computation of similarities. However, Hadoop is a framework based on hard disk, and constant reading and writing of data become the bottleneck in computation. Therefore, memory based framework Spark has become a prevalent solution for recommender systems. Panigrahi et al. used Alternating Least Square (ALS) on Spark and k-means to avoid the data sparsity and scalability of collaborative filtering algorithms [43]. Wijayanto and Winarko implemented multicriteria collaborative filtering using Spark framework [44]. The experiments' results showed that efficiency of algorithms improved with the number of nodes in Spark clusters. Therefore, in order

to obtain higher computing efficiency, it is necessary to use Spark in recommender systems.

2.6. The Contribution of Our Work. As mentioned before, various recommendation models have been suggested as powerful tools for movie recommendation. Previous practitioners and academic researchers focus on the improvement of the recommendation performance by using the combination of recommendation models. However, they ignored that, with the increase of users and items to recommend, the computational overhead has heavily increased. Therefore, this paper proposes a sentiment-enhanced movie recommendation framework based on Spark platform to meet the requirement of mobile services in aspects of high timeliness. In our method, both the content-based method and the collaborative filtering method are taken into consideration. Based on collaborative filtering method and content-based method, the preliminary output is optimized by the analysis of the effect from both positive and negative information. Finally, experiments are carried out to prove the performance of our proposed method.

3. A Sentiment-Enhanced Recommendation Framework

As mentioned before, this paper uses collaborative filtering and content-based hybrid recommender systems. Collaborative filtering and content-based approaches can compensate for the shortcomings of each other, thus ensuring the accuracy and stability of the recommender system. On the one hand, collaborative filtering can make up for the lack of personalization of content-based method; on the other hand, content-based method can make up for the flaw of collaborative filtering method whose scalability is relatively weak. In general, the hybrid recommendation method is first executed based on user data and movie data to achieve a preliminary recommendation list. Then sentiment analysis is implemented to optimize the preliminary list and get the final recommendation list. Furthermore, on the basis of the hybrid recommendation framework, this paper fully considers the efficiency of the recommender system. In the process of recommending movies, this paper focuses on the user's reviews on movies. Under the influence of the herding effect, users are inclined to choose goods or services that most people prefer. Therefore, compared to movies with many negative reviews, movies with more positive reviews will be given priority to be recommended to users. After optimization, final recommendation list is generated, as shown in Figure 1.

3.1. Data Collection. In this paper we use data derived from Douban movie (https://movie.douban.com/) to verify the validity of the model we proposed. Douban movie data can be divided into user data, movie data, and review data. User data and movie data are used as the input of collaborative filtering method, while review data are used as the material of content-based method. As input of the model, data need preprocessing, which includes data clean, data integration, and data transformation.

3.2. The Hybrid Recommendation Module. In our proposed method, hybrid recommendation method is basic to the generation of a preliminary recommendation list. To process the hybrid method on Spark, the following steps are needed.

Step 1 (collect user preferences and item representation). Collaborative filtering method is used to discover principles from users' behavior and preferences, so how to collect the user's preferences becomes the basis of the method. Users have lots of ways to provide their own preferences for the system, such as ratings and clicks. In our proposed method, users' ratings on movies are taken into consideration. We need to preprocess the data before we import the data into the collaborative filtering model. The core of the work is normalization and reducing noise. First, noise should be filtered out because the existence of noise will result in a decrease in the efficiency and effectiveness of the recommender system. Second, the input data need normalization. By normalizing data, the method can be made more accurate.

Through the above steps, we get a two-dimensional table, in which one dimension is the user list, and the other dimension is the movie list, while the value is the user's ratings for movies. The preference data is transformed into user-movie resilient distributed datasets (RDD), which can be processed by Spark. From user's behavior and preferences we can discover some disciplines to help the following recommendation.

Due to the high timeliness requirement of mobile services, we need to improve the efficiency of the calculation. In the process of computing user preferences, the data are stored in the memory of Spark. If the calculation steps of content-based recommendation are processed after the data are written to disks, unnecessary I/O will be carried out. As a result, we tend to read data into memory and compute user preferences for collaborative filtering method and item representation for content-based method simultaneously. In our proposed method, movies are represented by their genres, directors, and actors.

Step 2 (distributed process). In order to process the data in a distributed form, Spark platform calculated the total number of items each user prefers and the total number of items that any two users prefer at the same time. The two kinds of statistics can be distributed on the computing nodes of the Spark platform and the results are stored in the form of RDD, respectively.

Step 3 (find similar users). After getting the user's preferences by analyzing users' behavior, similar users and items can be calculated based on the users' preferences.

To find similar users, similarity between users should be calculated. In this paper, we employ Euclidean distance to measure the similarity. Therefore, the similarity between users u_x, u_y can be calculated by

$$\operatorname{sim}\left(u_x, u_y\right) = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - y_i\right)^2}{n}}, \tag{1}$$

where x_i, y_i represent the ratings from u_x, u_y on movie i.

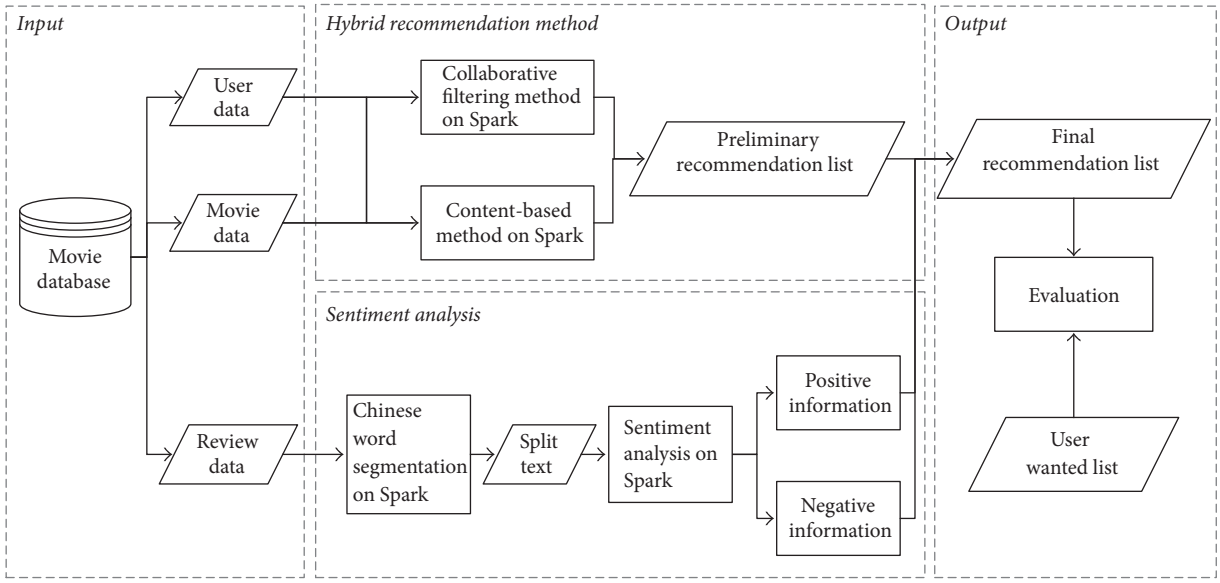


Figure 1: A sentiment-enhanced hybrid recommendation framework.

Step 4 (calculate and recommend). In the previous steps, all users can be ranked according to the value $\mathrm{sim}(u_x, u_y)$. In order to recommend movies to user u_x, top K most similar users are selected. Then according to their similarities and preferences for movies, a list of recommended movies is calculated to be supplied for user u_x. Moreover, the similarities between the preference of user u_x and item representation vectors are also taken into consideration. Movies that are not suitable for the user u_x will be removed from the list. Then the list is the preliminary recommendation list to be used as the foundation of our proposed method. We calculate scores derived from the two recommendation methods.

$$\begin{aligned} \mathrm{Score}_{\mathrm{CF},m} &= \sum_i \mathrm{sim}\left(u_t, u_i\right) R_{i,m} \\ \mathrm{Score}_{\mathrm{CB},m} &= R_{t,m} \sum_j \mathrm{sim}\left(m, m_{j,t}\right), \end{aligned} \tag{2}$$

where $\mathrm{Score}_{\mathrm{CF},m}$ represents the score of movie m in collaborative filtering. $\mathrm{sim}(u_t, u_i)$ denotes the similarity between user u_t and candidate user u_i. $R_{i,m}$ is the rating from candidate user u_i on movie m. $\mathrm{Score}_{\mathrm{CB},m}$ represents the score of movie m in content-based recommendation method. $\mathrm{sim}(m, m_{j,t})$ denotes the similarity between movie m and movies which user u_t have already watched. $R_{t,m}$ is the rating from user u_t on movie m.

3.3. The Sentiment-Based Recommendation Module. First, the algorithm will encounter text information that cannot be used directly. Therefore, text mining is introduced to extract information hidden in the text data. From the point of text processing involved in this article, there is no association between different reviews, so the data can be distributed directly without special treatment.

3.3.1. Chinese Word Segmentation. Text mining is used to extract useful information from text data [45, 46]. Due to the complexity of text data, researchers have invested great efforts to seek solutions for computers to understand the meaning of text [47]. Accordingly, some methods and changes must be done to process text data. First, we employ Chinese word segmentation to solve the problem. The tool used for Chinese word segmentation is ICTCLAS [48].

The movie reviews appear in the form of long sentences in different structures. Nevertheless, in one sentence, the main information of reviews exists in several words [49]. Hence, a few key words instead of the whole sentence should be analyzed. Chinese word segmentation is the basis of text mining in Chinese. For a Chinese sentence, Chinese word segmentation is the basis for computers to recognize meanings of text [50]. Unlike English and other languages, there is no space in Chinese as a natural separator [51]. At the lexical level, Chinese word segmentation is more complicated than English word segmentation. Different segmentation may lead to different understanding of Chinese. In this paper, the Chinese reviews of movies are divided into Chinese word sequences. At the same time, stop words are excluded to avoid their negative impact on the following sentiment analysis. After the Chinese word segmentation, the rest of the words are more relevant to our study.

3.3.2. Sentiment Analysis. After the Chinese word segmentation, we analyze the result of segmentation by sentiment analysis. Finally the review is expressed as a vector space model (VSM). The VSM assumes the words that make up the text are independent of each other, so that the text can be represented by these words, which provides the basis for the representation of the mathematical model. The expression of text as a VSM can make the text representation and processing convenient. The text category is only related to

specific words contained in the text and its frequency in the text. The review D can be expressed as a vector $D = \{(t_1, w_1), (t_2, w_2) \cdots (t_n, w_n)\}$. t_i represents the ith word in the review. w_i represents the weight of t_i. In this paper, we use term frequency-inverse document frequency (TF-iDF) value as feature weights.

After the vector space representation of the movie reviews are obtained, the sentiment analysis based on the lexicon can be carried out smoothly. We first classify the reviews into positive and negative parts according to the sentiment lexicon. The lexicon is built according to the field of movies. Words such as "good" and "wonderful" in reviews indicate that the user had a positive impression of the movie. If most users have positive evaluation on the movie, the movie should be deemed as a priori one to be recommended to users who have not watched it.

After analyzing and processing the sentiment words in the movie reviews and the sentiment lexicon of the corresponding categories of movie reviews, the sentiment value H is calculated, and H_m represents the sentiment value of review m.

$$H_m = \sum_{n=1}^{l} W_l w_l, \tag{3}$$

where W_l represents the weight of the words in the lexicon of the corresponding movie category and w_l represents the weight of the words in the vector space representation.

3.4. Ranking and Recommendation. The preliminary recommendation list based on hybrid recommendation method contains movies ranked by their scores. The scores are derived from the calculation of the collaborative filtering and content-based recommendation method, as shown in the following formula:

$$\text{Score}_{\text{hybrid},m} = \text{Score}_{\text{CF},m} + \text{Score}_{\text{CB},m}, \tag{4}$$

where $\text{Score}_{\text{hybrid},m}$ represents the score of movie m in the hybrid recommendation system.

Sentiment analysis will optimize the preliminary recommendation list. The sentiment score will be added to the score of the movie. Therefore, the score of each film is as follows:

$$\text{Score}_{\text{final},m} = W_{\text{hybrid}} \text{Score}_{\text{hybrid},m} + W_{\text{SA}} \text{Score}_{\text{SA},m}, \tag{5}$$

where W_{hybrid} and W_{SA} represent the weights of two recommendation methods and $\text{Score}_{\text{SA},m}$ represents the score of movie m derived from sentiment analysis. $\text{Score}_{\text{SA},m}$ is the sum of all H_m of reviews for the movie.

Final recommendation list is generated according to the new score. The wanted list is a group of movies with no order. To adapt to this situation, the final recommendation list will be present with no order. Therefore, in order to select enough appropriate movies, more movies are selected by hybrid recommendation method and some are discarded by final scores.

The recommender system displays the optimized list of recommendations to users. Douban movie users have "wanted list" which lists movies that users want to see but have not seen. Therefore, this paper uses the "wanted list" to evaluate the proposed model.

Table 1: The confusion matrix.

	Recommendation list	
	In the list	Not in the list
Wanted list		
In the list	TP	FN
Not in the list	FP	TN

4. Empirical Analysis

4.1. Data Description. The data used in this paper is real-world data derived from Douban movie, a website that provide users with information of movies. Users can make reviews on each movie they have seen. The user-generated reviews are shown to other users who have desire to see the movie.

We ultimately get 12253 available items in the data, and each item represents a movie. As a whole, there are 6179857 reviews of these movies from 205754 users. On average, there are about 504 reviews for each movie and every user makes 30 reviews. Moreover, users' ratings on these movies are also obtained from the website.

4.2. Evaluation Criterion. To evaluate the performance of our model, four criteria are used to evaluate the results. The criteria are derived from the confusion matrix, as shown in Table 1.

Precision and recall are contradictory to some extent, so we employ F-measure. F-measure is the weighted harmonic average of precision and recall, which can better measure the performance of the model in a more comprehensive prospect.

$$\begin{aligned} \text{TP rate} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ F1 &= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \\ \text{FP rate} &= \frac{\text{FP}}{\text{FP} + \text{TN}}. \end{aligned} \tag{6}$$

4.3. Experimental Results. The output of sentiment analysis applied on the reviews of the movies is affiliated to the evaluation of preliminary recommendation list. Sentiment analysis can optimize the candidate movie list. Therefore, the combination of collaborative filtering and content-based method with sentiment analysis makes our model performs better. For comparison, we also evaluate some recommendation method. The experimental results are shown in Table 2 and Figure 2. Our model performs better than basic recommendation in terms of TP rate, which means that our model is stronger in the ability to identify appropriate movies. CF is

Table 2: The performance of recommendation models.

	TP rate	FP rate	Precision	$F1$
CF + CB	0.645	0.355	0.531	0.582
CF + CB + SA	0.761	0.239	0.782	0.771

Table 3: The running time of the hybrid recommender system on Spark.

Number of nodes	Full data running time on Spark/seconds	Half data running time on Spark/seconds
(1)	463	246
(2)	276	151
(3)	197	114
(4)	153	92
(5)	101	63
(6)	86	56
(7)	75	48
(8)	65	44
(9)	58	41

short for collaborative filtering. CB represents content-based method, and SA is short for sentiment analysis.

We also compared running time on different number of nodes and different amount of data. The experimental results are shown in Table 3 and Figure 3.

First, as the number of nodes in the computational cluster increases, the computational efficiency of Spark is increasing, and the corresponding experimental result shows that the running time decreases. Second, when our model is applied in larger data, the speedup of computational efficiency is better. The results show that our proposed method performs well both in accuracy and efficiency. On the one hand, it can help merchants avoid customer churn due to delayed information and recommendation provided for mobile services users. On the other hand, it can provide help for improving the timeliness satisfaction of mobile services users.

5. Conclusions and Future Work

Mobile recommender system requires both accuracy and timeliness. In this paper, a movie recommendation framework based on hybrid recommendation and sentiment analysis is proposed to improve the accuracy of recommender systems. Furthermore, Spark is used to improve the timeliness of the system. Our proposed method makes it convenient and fast for users to obtain useful movie suggestions. Movie recommendation is a comprehensive task which involves various kinds of users and various kinds of movies. Considering the useful information hidden in reviews posted by users, collaborative filtering is considered to be the most popular and widely deployed technique in recommender system. Moreover, due to the characteristics of movie recommendation, the user watching history is very important, so we add content-based recommendation method to collaborative filtering to compose a hybrid recommender system. Moreover, it is

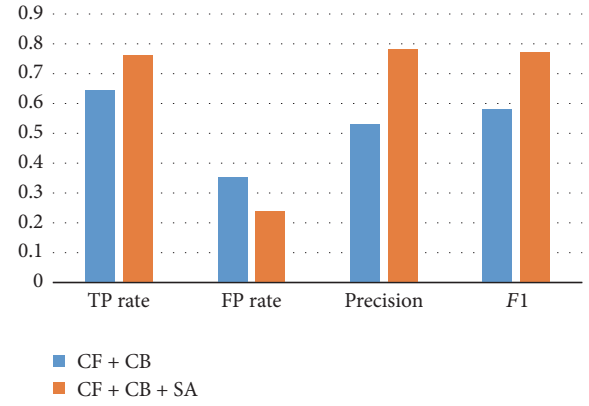


Figure 2: The performance of recommendation models.

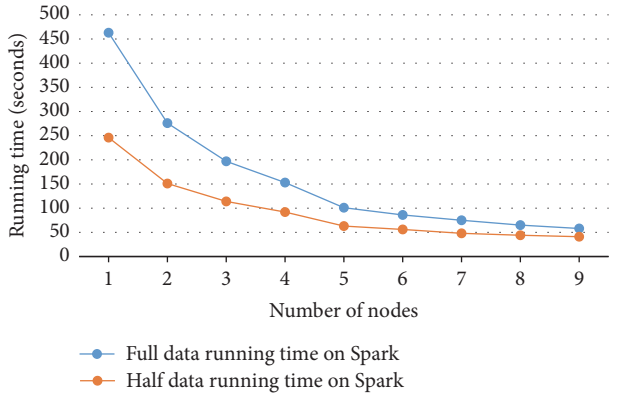


Figure 3: The running time of the hybrid recommender system on Spark.

better to consider the sentiment of positive and negative information during the analysis of recommender system. In general, people tend to think that positive reviews have a positive impact and negative reviews have negative effects. Sentiment analysis will help us improve the accuracy of recommendation results. Furthermore, as we illustrated in our experimental results, it is necessary to employ distributed system to solve the scalability and timeliness of recommender system.

The proposed framework can be improved in several aspects. First, this method can be verified in more data sets. Different data can be used by different sentiment analysis, so the model can be tuned to accommodate more situations. Second, in the analysis process of the sentiment analysis, different kinds of subjective ideas are involved inevitably, which implements adverse effects on the results. Therefore, future work will focus on the eliminating of individual characteristics hidden in the text description from users.

Disclosure

Mingming Wang and Wei Xu are the corresponding authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grants nos. 71301163, 71771212), Humanities and Social Sciences Foundation of the Ministry of Education (nos. 14YJA630075, 15YJA630068), the Fundamental Research Funds for the Central Universities, the Research Funds of Renmin University of China (no. 15XNLQ08), and the Outstanding Innovative Talents Cultivation Funded Programs 2017 of Renmin University of China.

References

[1] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," in *Proceedings of the 1998 15th National Conference on Artificial Intelligence, AAAI*, pp. 714–720, July 1998.

[2] W. Xu, J. Wang, Z. Zhao, C. Sun, and J. Ma, "A Novel Intelligence Recommendation Model for Insurance Products with Consumer Segmentation," *Journal of Systems Science and Information*, vol. 2, no. 1, pp. 16–28, 2014.

[3] Y. Xu, X. Guo, J. Hao, J. Ma, R. Y. K. Lau, and W. Xu, "Combining social network and semantic concept analysis for personalized academic researcher recommendation," *Decision Support Systems*, vol. 54, no. 1, pp. 564–573, 2012.

[4] D. Guo, Z. Zhao, W. Xu et al., "How to find a comfortable bus route - Towards personalized information recommendation services," *Data Science Journal*, vol. 14, article no. 14, 2015.

[5] D. Guo, Y. Zhu, W. Xu, S. Shang, and Z. Ding, "How to find appropriate automobile exhibition halls: Towards a personalized recommendation service for auto show," *Neurocomputing*, vol. 213, pp. 95–101, 2016.

[6] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang, "Collaborative filtering meets mobile recommendation: a user-centered approach," in *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, pp. 236–241, 2010.

[7] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Towards mobile intelligence: learning from GPS history data for collaborative recommendation," *Artificial Intelligence*, vol. 184/185, pp. 17–37, 2012.

[8] M.-H. Park, J.-H. Hong, and S.-B. Cho, "Location-based recommendation system using Bayesian user's preference model in mobile devices," in *Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing*, pp. 1130–1139, 2007.

[9] C. Ono, M. Kurokawa, Y. Motomura, and H. Asoh, "A context-aware movie preference model using a bayesian network for recommendation and promotion," in *Proceedings of 11th International Conference on User Modeling*, pp. 247–257, 2007.

[10] M. Szomszor, C. Cattuto, H. Alani et al., "Folksonomies, the semantic web, and movie recommendation," in *Proceedings of 4th European Semantic Web Conference*, pp. 1–14, 2007.

[11] T. De Pessemier, T. Deryckere, and L. Martens, "Context aware recommendations for user-generated content on a social network site," in *Proceedings of the EuroITV'09 - 7th European Conference on European Interactive Television Conference*, pp. 133–136, Belgium, June 2009.

[12] J. L. Herlocker, J. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, pp. 230–237, Berkeley, Calif, USA, August 1999.

[13] Y. Koren, "Collaborative filtering with temporal dynamics," *Communications of the ACM*, vol. 53, no. 4, pp. 89–97, 2010.

[14] T. Hofmann, "Collaborative filtering via gaussian probabilistic latent semantic analysis," in *Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Informaion Retrieval (SIGIR)*, pp. 259–266, Toronto, Canada, 2003.

[15] Y. Zhang, D. Zhang, M. M. Hassan, A. Alamri, and L. Peng, "CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies," *Mobile Networks and Applications*, vol. 20, no. 3, pp. 348–355, 2015.

[16] B. Marlin, "Modeling user rating profiles for collaborative filtering," in *Proceedings of the 16th International Conference on Neural Information Processing Systems*, pp. 627–634, 2003.

[17] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pp. 1257–1264, 2007.

[18] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proceedings of the 24th International Conference on Machine learning (ICML '07)*, vol. 227, pp. 791–798, Corvallis, Oregon, June 2007.

[19] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*, pp. 426–434, New York, NY, USA, August 2008.

[20] S. Rendle, "Factorization machines," in *Proceedings of the 10th IEEE International Conference on Data Mining, ICDM 2010*, pp. 995–1000, Australia, December 2010.

[21] Y. Zhen, W.-J. Li, and D.-Y. Yeung, "TagiCoFi: Tag informed collaborative filtering," in *Proceedings of the 3rd ACM Conference on Recommender Systems, RecSys'09*, pp. 69–76, USA, October 2009.

[22] G. Lekakos and P. Caravelas, "A hybrid approach for movie recommendation," *Multimedia Tools and Applications*, vol. 36, no. 1-2, pp. 55–70, 2008.

[23] Y. Zhang, Z. Tu, and Q. Wang, "TempoRec: Temporal-Topic Based Recommender for Social Network Services," *Mobile Networks and Applications*, vol. 22, pp. 1182–1191, 2017.

[24] S. Debnath, N. Ganguly, and P. Mitra, "Feature weighting in content based recommendation system using social network analysis," in *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*, pp. 1041-1042, Beijing, China, April 2008.

[25] M. Nazim Uddin, J. Shrestha, and G.-S. Jo, "Enhanced content-based filtering using diverse collaborative prediction for movie recommendation," in *Proceedings of the 2009 1st Asian Conference on Intelligent Information and Database Systems, ACIIDS 2009*, pp. 132–137, Viet Nam, April 2009.

[26] A. Gunawardana and C. Meek, "A unified approach to building hybrid recommender systems," in *Proceedings of the 3rd ACM Conference on Recommender Systems, RecSys'09*, pp. 117–124, USA, October 2009.

[27] K. Soni, R. Goyal, B. Vadera, and S. More, "A Three Way Hybrid Movie Recommendation Syste," *International Journal of Computer Applications*, vol. 160, no. 9, pp. 29–32, 2017.

[28] G. Ling, M. R. Lyu, and I. King, "Ratings meet reviews, a combined approach to recommend," in *Proceedings of the 8th ACM Conference on Recommender Systems, RecSys 2014*, pp. 105–112, USA, October 2014.

[29] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "IDoctor: personalized and professionalized medical recommendations

based on hybrid matrix factorization," *Future Generation Computer Systems*, vol. 66, pp. 30–35, 2017.

[30] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing—Volume 10 (EMNLP '02)*, pp. 79–86, Association for Computational Linguistics, Stroudsburg, Pa, USA, July 2002.

[31] P. D. Turney, "Thumbs up or thumbs down?" in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417–424, Philadelphia, Pennsylvania, July 2002.

[32] J. R. Priester and R. E. Petty, "The Gradual Threshold Model of Ambivalence: Relating the Positive and Negative Bases of Attitudes to Subjective Ambivalence," *Journal of Personality and Social Psychology*, vol. 71, no. 3, pp. 431–449, 1996.

[33] H. Li, J. Cui, B. Shen, and J. Ma, "An intelligent movie recommendation system through group-level sentiment analysis in microblogs," *Neurocomputing*, vol. 210, pp. 164–173, 2016.

[34] J. Sun, G. Wang, X. Cheng, and Y. Fu, "Mining affective text to improve social media item recommendation," *Information Processing & Management*, vol. 51, no. 4, pp. 444–457, 2015.

[35] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang, "Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS)," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, pp. 193–202, ACM, New York, NY, USA, August 2014.

[36] Y. Zhang, "GroRec: A Group-Centric Intelligent Recommender System Integrating Social, Mobile and Big Data Technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[37] D. Liu, W. Xu, W. Du, and F. Wang, "How to choose appropriate experts for peer review: An intelligent recommendation method in a big data context," *Data Science Journal*, vol. 14, article no. 16, 2015.

[38] W. Xu, J. Sun, J. Ma, and W. Du, "A personalized information recommendation system for RD project opportunity finding in big data contexts," *Journal of Network and Computer Applications*, vol. 59, pp. 362–369, 2016.

[39] Y. Zhou, D. Wilkinson, R. Schreiber, and R. Pan, "Large-scale parallel collaborative filtering for the Netflix Prize," in *Proceedings of the 4th International Conference on Algorithmic Aspects in Information and Management*, pp. 337–348, 2008.

[40] Z.-D. Zhao and M.-S. Shang, "User-based collaborative-filtering recommendation algorithms on hadoop," in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, WKDD 2010*, pp. 478–481, Thailand, January 2010.

[41] J. Sun, W. Xu, J. Ma, and J. Sun, "Leverage RAF to find domain experts on research social network services: A big data analytics methodology with MapReduce framework," *International Journal of Production Economics*, vol. 165, pp. 185–193, 2015.

[42] J. Jiang, J. Lu, G. Zhang, and G. Long, "Scaling-up item-based collaborative filtering recommendation algorithm based on Hadoop," in *Proceedings of the 7th IEEE World Congress on Services*, pp. 490–497, IEEE, Washington, DC, USA, July 2011.

[43] S. Panigrahi, R. K. Lenka, and A. Stitipragyan, "A Hybrid Distributed Collaborative Filtering Recommender Engine Using Apache Spark," in *Proceedings of the 7th International Conference on Ambient Systems, Networks and Technologies, ANT 2016 and the 6th International Conference on Sustainable Energy Information Technology, SEIT 2016*, pp. 1000–1006, Spain, May 2016.

[44] A. Wijayanto and E. Winarko, "Implementation of multi-criteria collaborative filtering on cluster using Apache Spark," in *Proceedings of the 2nd International Conference on Science and Technology-Computer, ICST 2016*, pp. 177–181, Indonesia, October 2016.

[45] R. Feldman and I. Dagan, "Knowledge discovery in textual databases (KDT," in *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, pp. 112–117, 1995.

[46] A. H. Tan, "Text mining: the state of the art and the challenges," in *Proceedings of the Pacific Asia Conference on Knowledge Discovery and Data Mining PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, pp. 65–70, 1999.

[47] A. Hotho, A. Nürnberger, and G. Paaß, "A brief survey of text mining," *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*, vol. 20, pp. 19–62, 2005.

[48] H.-P. Zhang, H.-K. Yu, D.-Y. Xiong, and Q. Liu, "HHMM-based Chinese lexical analyzer ICTCLAS," in *Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing (SIGHAN '03)*, pp. 184–187, Sapporo, Japan, July 2003.

[49] Y. Wang and W. Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud," *Decision Support Systems*, vol. 105, pp. 87–95, 2018.

[50] F. Wu, Y. Huang, Y. Song, and S. Liu, "Towards building a high-quality microblog-specific Chinese sentiment lexicon," *Decision Support Systems*, vol. 87, pp. 39–49, 2016.

[51] Y. Wang, W. Xu, and H. Jiang, "Using text mining and clustering to group research proposals for research project selection," in *Proceedings of the 48th Annual Hawaii International Conference on System Sciences, HICSS 2015*, pp. 1256–1263, USA, January 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 5048419, 7 pages
https://doi.org/10.1155/2018/5048419

Research Article

Sampling Adaptive Learning Algorithm for Mobile Blind Source Separation

Jingwen Huang[1] and Jianshan Sun[2]

[1]*Beijing University of Chemical Technology, Beijing 100029, China*
[2]*School of Management, Hefei University of Technology, Hefei 230009, China*

Correspondence should be addressed to Jianshan Sun; sunjs9413@hfut.edu.cn

Received 23 November 2017; Accepted 26 December 2017; Published 18 March 2018

Academic Editor: Yin Zhang

Learning rate plays an important role in separating a set of mixed signals through the training of an unmixing matrix, to recover an approximation of the source signals in blind source separation (BSS). To improve the algorithm in speed and exactness, a sampling adaptive learning algorithm is proposed to calculate the adaptive learning rate in a sampling way. The connection for the sampled optimal points is described through a smoothing equation. The simulation result shows that the performance of the proposed algorithm has similar Mean Square Error (MSE) to that of adaptive learning algorithm but is less time consuming.

1. Introduction

With the fast development of the information and computation technologies, the big data analysis and cognitive computing have been widely used in many research areas such as medical treatment [1], transportation [2], and wireless communication [3, 4]. Blind Source Separation (BSS) is a popular research topic in the area of wireless communication. With the fast development of mobile computing, BSS has been widely used in the mobile signal analysis. BSS is an integration of artificial neural network, statistical signal processing, and information theory. The core of BSS is its ability to extract independent components from an observed mixture signal, without requiring a prior knowledge. Such flexibility has made BSS popular in many applications [5–7] especially in mobile intelligence [8, 9].

Artificial neural network based Independent Component Analysis (ICA) is the widely used method in BSS, because it provides powerful tools to capture the structure in data by learning. Based on this theory, Natural Gradient Algorithm (NGA) is employed to find the appropriate coefficient vector of artificial neural network [10]. Nonholonomic Natural Gradient Algorithm (NNGA) is addressed and applied in the BSS [11, 12]. In its application, learning rate for training coefficient vector plays an important role on the performance of the algorithm, which has relationship with not only the update times but also the speed of convergence. This attracts many researchers' attention to the learning algorithms [13, 14].

Most well-known traditional learning algorithms assume that the learning rate is a small positive constant. Inappropriate constant will lead to relative slow convergence speed or big steady state error. There are lots of studies on the learning rate which aim at the better performance and higher convergence speed. von Hoff and Lindgren [15] developed adaptive step size control algorithm for gradient-based BSS. They used the coefficients of the estimating function to provide an appropriate "measure of error" and serve as the basis for a self-adjusting time-varying step-size. Hai proposed a conjugate gradient procedure for second-order gradient-based BSS [16]. The second-order gradient-based BSS was formulated as a constrained optimization problem. A conjugate gradient procedure with the step size derived optimally at each iteration was proposed to solve the optimization problem. In these algorithms, the step size is updated in iteration, whose value is adjusted according to the time-varying dynamics of the signals. These approaches lead to better performance. The real time search for the step size, however, requires more online calculations as well as recursion, resulting in an increase in computational complexity. On the other hand, in the recursion for optimal step size, there is still

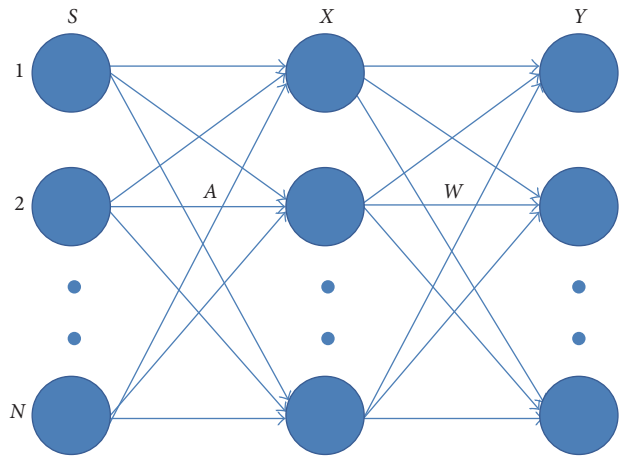


Figure 1: The system model.

a constant left to be estimated, which leads to an endless loop.

The objective of this paper is to find appropriate learning algorithm, to provide better performance as well as less computation time. The proposed sampling learning rate is based on adaptive learning algorithm, which only calculates and samples a few appropriate points. These selected points are connected by the proposed normalized smooth equation.

In the following, we first review the principle of blind signal separation. Then, we discuss the adaptive learning algorithm and propose the sampling adaptive learning algorithm. Finally, we present two typical examples in mobile voice signal. Different constant learning rates are compared firstly to analyze the relationship between the convergence speed and steady state error. Then the comparison between the adaptive learning algorithm and the sampling adaptive learning algorithm is made, illustrating that the proposed algorithm has similar Mean Square Error (MSE) to that of the adaptive learning algorithm but consumes less computational time.

2. Blind Signal Separation

The model considered in this paper is described by Figure 1. A set of individual source signals $\mathbf{s}(k)$ is mixed with A matrix to produce a set of mixed signals $\mathbf{x}(k)$.

$$\mathbf{x}(k) = A\mathbf{s}(k), \tag{1}$$

where $A = [a_{ij}] \in \mathbb{R}$ is the an unknown mixing matrix independent of time, $\mathbf{s}(k) = [s_1(k), s_2(k), \ldots, s_N(k)]^T$ is the vector of source signals, and $\mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_N(k)]^T$ is the vector containing the observed signals.

BSS separates the mixed signals, through the determination of an unmixing matrix $W(k)$ to recover the approximation of the original signals. The recovered output signal is described by

$$\mathbf{y}(k) = W(k)\mathbf{x}(k), \tag{2}$$

where W is the $N \times N$ matrix to be adjusted such that $WA = I$. With the NGA, the matrix W is updated from W_k at time k to W_{k+1} by using the following adaptation rule.

$$W(k+1) = W(k) + \partial(k)\left[I - f(y(k))\, y^T(k)\right] W(k), \tag{3}$$

where $\partial(k)$ is the learning rate. $f(y)$ is the score function defined by

$$f_i(y_i) = -\frac{p_i(y_i)'}{p_i(y_i)}, \tag{4}$$

where $p_i(y_i)$ is the probability density function of the ith source signal. It is assumed that the source signals are zero mean. Hence, we have

$$E(s_i(t)) = 0, \tag{5}$$

where E denotes the expectation.

According to the reference [17], $f(y)$ can be set to be $f(y) = y^3$ when $\mathbf{s}(k)$ is the sub-Gaussian signal; $f(y) = \tan(y)$ when $\mathbf{s}(k)$ is the super-Gaussian signal. Most mobile voice signals are super-Gaussian signals and most mobile image signals are sub-Gaussian signals. The function is accordingly selected based on the mobile signal type.

The learning rate is a very important factor for the performance of BSS in controlling the magnitudes of the updates of the estimated parameters. It can be constant or variable. The constant learning rate means that the adaptation in (3) is based on fixed step-size parameter. Because the step size is proportional to convergence speed and the steady state error, the big step size leads to high convergence speed and big steady state error. However, what we expect is the high convergence speed and small steady state error. In conclusion, the main problem of the constant learning rate focuses on the incompatibility between the convergence speed and the steady state error [18, 19]. Therefore, many researchers were aiming at adaptiveness by using the variable step-size approaches. We will discuss the adaptive learning algorithm and propose a sampling adaptive learning algorithm in the following section.

3. Sampling Adaptive Learning Algorithm

3.1. Adaptive Learning Algorithm for BSS. The idea of adaptively changing the step size of the learning rate is called adaptive learning algorithm. The adaptive learning algorithm can balance the convergence speed and the steady error performance. In this case, we discuss the adaptive learning algorithm which updates the learning rate step size through the estimate function. Although the distance between the estimated parameter and its optimal value is not directly available to control the step size, evaluation function can be developed to estimate the distance so that the step size can be determined by the following recursion:

$$\partial(k+1) = \partial(k) + \beta\overline{\varphi}(H_k), \tag{6}$$

Table 1: Performance of CLA and ALA.

	Convergence time (iterative time)	Mean steady state error	Computation time
CLA	681	2.1	55.6 s
ALA	482	1.9	96.6 s

where β is a constant and H_k is an estimate function. φ is the evaluation function. In this adaptive learning algorithm, the estimate function is set to be

$$H_k = I - f_k(y)\, y_k^T. \tag{7}$$

A smoothed version of the evaluation function, denoted by $\widehat{H}_{ijk}$, can be set to be

$$\begin{aligned}\widehat{H}_{ijk} &= E\left(H_{ijk}\right) \\ &= -E\left(f'\left(s_i\right)\right) E\left(s_j^2\right) W_{ijk}A - E\left(f\left(s_i\right) s_i\right) W_{jik}A\end{aligned} \tag{8}$$

and the evaluation function $\overline{\varphi}(H_k)$ is set to be

$$\overline{\varphi}\left(H_k\right) = \max\left(\left|\widehat{H}_{ijk}\right|\right). \tag{9}$$

This idea is similar to Least Mean Square Adaptive Filter (LMS) and Reinforcement Learning (RL). It can be concluded from (6)–(9) that the step size depends on the evaluation of the estimate function. The principle of this adaptive algorithm implies that the step size is small when the errors are small and the step size is large when the errors are large. For time-invariant systems the step size systematically decreases when the learning rate is close to their optimum.

However, we must notice that, in the process to obtain a better learning rate, the adaptive step-size control algorithm actually introduces another recursion. In this recursion shown in (6), there is still an unknown constant β to be determined. If we want to find the optimal β, we also need to find another recursion like (6), which leads to endless iteration.

On the other side, the recursion as described in (6) introduces another calculation recycle which adds additional computation. In some case, calculation with adaptive learning rate even consumes more time than the calculation with constant learning rate. Figure 2 is the ideal adaptive step size curve for noise-free signal. Every point on the curve is calculated with (6)–(9). We did the simulation for BSS based on the adaptive step-size and compared it with the selected ideal constant step size. Table 1 shows the performance index resulting from the application of the two methods and illustrates that the convergence speed of ALA is considerably higher when the adaptive step-size algorithm is employed. The steady state error of these two algorithms is similar when the random possibility is considered. As for the computation time, it is obvious that ALA consumes more. We can conclude from the analysis that although the adaptive algorithm balances the convergence speed and the steady state error, it consumes more computation time as shown in Table 1.

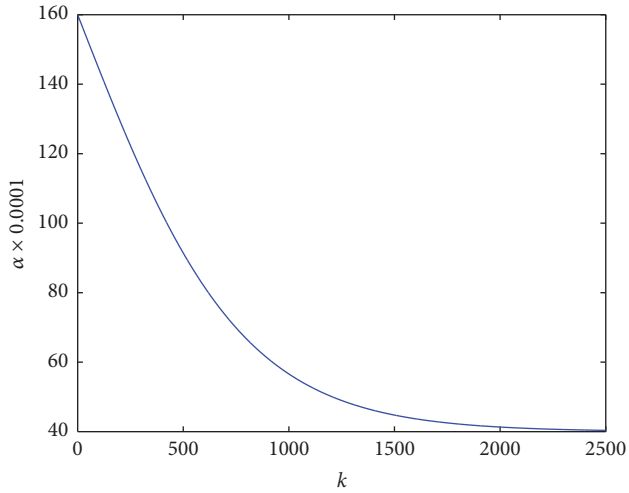


Figure 2: The ideal adaptive step size curve for noise-free signal.

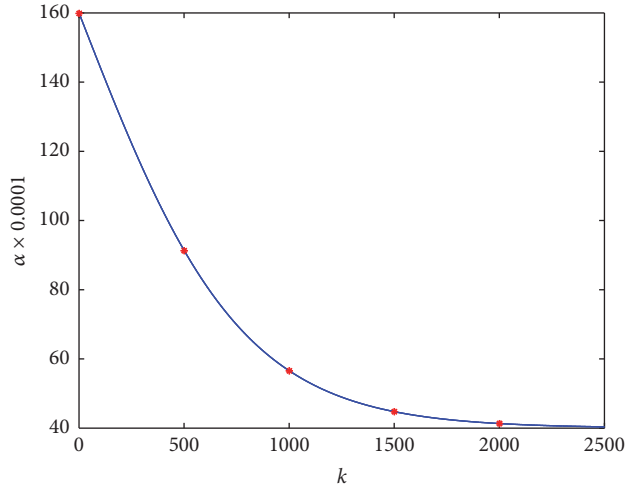


Figure 3: Effect of sample learning rate on updating the matrix W.

3.2. Adaptive Learning Algorithm Based on Sampling. An alternative way to use the adaptive strategy is to implement the adaptive step size on sampling. In this method, only several points need the adaptive calculation. For training the W matrix, it is not necessary to have the variable step size all the time. As shown in Figure 3, several sampling points are enough for keeping W updated. Let ΔK be the sampling interval; the time variable k' becomes

$$k' = \left[\frac{k}{\Delta K}\right], \tag{10}$$

where [ ] means rounding.

Then the learning rate can be represented as

$$\partial\left(k'+1\right) = \partial\left(k'\right) + \beta\overline{\varphi}\left(H_{k'}\right). \tag{11}$$

Equations ((7)–(9)) therefore become

$$\begin{aligned} H_{k'} &= I - f_{k'}(y)\, y_{k'}^{T} \\ \widehat{H}_{ijk'} &= E\left(H_{ijk'}\right) \\ &= -E\left(f'(s_i)\right) E\left(s_j^2\right) W_{ijk'} A \\ &\quad - E\left(f(s_i)\, s_i\right) W_{jik'} A \\ \overline{\varphi}\left(H_{k'}\right) &= \max\left(\left|\widehat{H}_{ijk'}\right|\right). \end{aligned} \tag{12}$$

The new equations reduce the times of iteration by dividing the sampling interval ΔK. We could choose the ΔK according to the required accuracy and speed. We also provide the connection for the sampled optimal points, which smooth the curve between two optimal points. Based on this analysis, the learning rate between two optimal points can be expressed as

$$\partial(k) = (1 - \tanh(\varsigma k))\, \partial_{1s} + \tanh(\varsigma k)\, \partial_{2s}, \tag{13}$$

where ∂_{1s} and ∂_{2s} are two sampling points which can be obtained at times k' and $k' + 1$ with sampling interval. When $\varsigma k = 5.3$, $\tanh(\varsigma k) \approx 1$ can be added as the condition used to determine where to switch the step size. In (13), k is the only variable that keeps changing. This algorithm will not bring another recursion for the system but will still have the optimal value choice.

For the convenience of application, the normalized form for $\partial(k)$ is given by

$$\begin{aligned} \partial(k) &= \left(1 - \tanh\left(100\varsigma_N \frac{k - ks_1}{k_{s1} - k_{s2}}\right)\right) \partial_{1s} \\ &\quad + \tanh\left(100\varsigma_N \frac{k - ks_1}{k_{s1} - k_{s2}}\right) \partial_{2s}, \end{aligned} \tag{14}$$

where ς_N is set to be 0.053. k_{s1} and k_{s2} are the first and second sampling points, respectively.

4. Experiment and Result

4.1. Case 1. To test the algorithm, five sub-Gaussian source signals commonly studied in the mobile system are employed. The source signals are

$$\begin{aligned} s(1) &= \sin(2\pi 800t) \\ s(2) &= \cos(2\pi 500t) \\ s(3) &= \sin(2\pi 100t + 6\cos(2\pi 40t)) \\ s(4) &= \sin(2\pi 50t) \\ s(5) &= \text{rand}(1, \text{length}(t)). \end{aligned} \tag{15}$$

These source signals are shown in Figure 5. A mixing coefficient was assigned to be independent white Gaussian noise, which is with zero mean. These sources were mixed producing mixtures as shown in Figure 6. The mixtures were separated using fixed learning rate, variable learning rate, and sampled learning rate which are named Constant Learning Algorithm (CLA), Adaptive Learning Algorithm (ALA), and sampling adaptive learning algorithm (SALA).

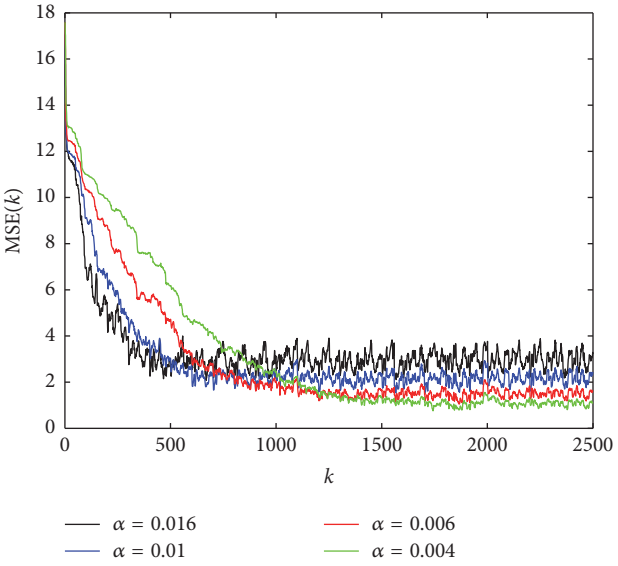


Figure 4: Mean Square Error (MSE) for different fixed learning rate.

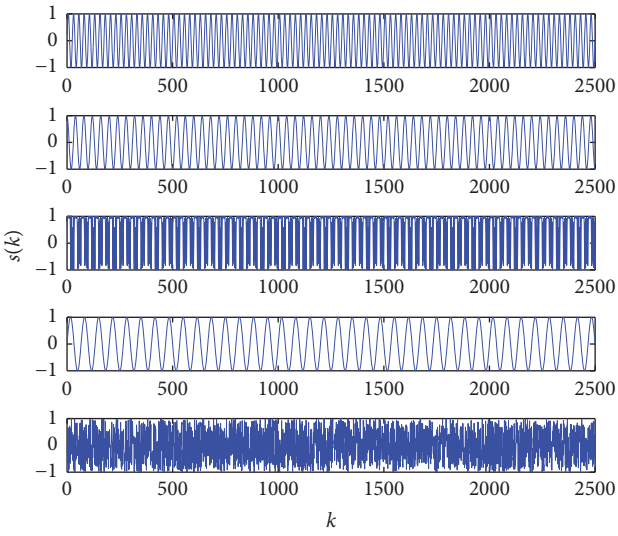


Figure 5: The raw source signals.

Firstly, we employ four fixed learning rate $\partial = [0.016, 0.010, 0.006, 0.004]$ to check the performance. Figure 4 depicts the Mean Square Error (MSE) for these four fixed learning rates. In this case the MSE [20] is defined as

$$\begin{aligned} \text{MSE}(k) = \sum_{i=1}^{n} \Bigg\{ &\left(\sum_{k=1}^{n} \frac{|W_{ik}A|}{\max_j |W_{ij}A|} - 1 \right) \\ &+ \left(\sum_{k=1}^{n} \frac{|W_{ki}A|}{\max_j |W_{ji}A|} - 1 \right) \Bigg\}. \end{aligned} \tag{16}$$

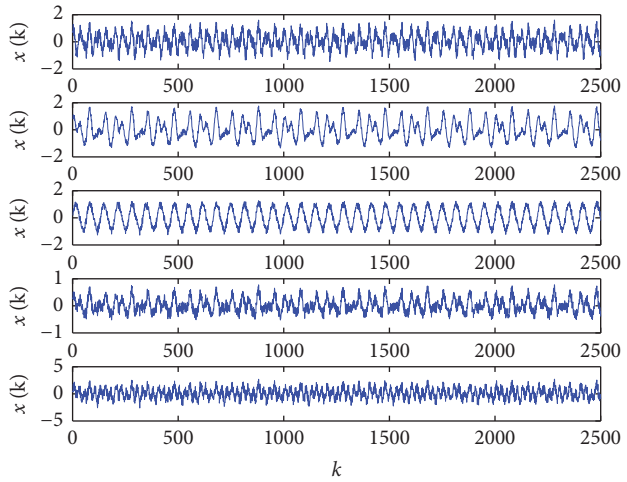


Figure 6: The mixed signals.

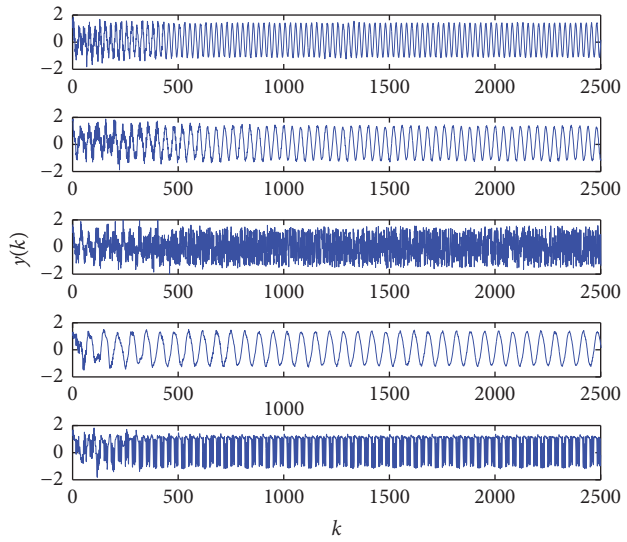


Figure 7: The separated signals.

As can be seen from Figure 4, MSE varied according to different fixed learning rate. The bigger learning rate has fast convergence speed at the beginning while the smaller learning rate leads to better stability at the end.

Adaptive Learning Algorithm (ALA) can balance the requirement of convergence speed and steady state error. However, ALA requires more computation time in iteration. To solve this problem, we use the sampling adaptive learning algorithm (SALA), by using the sampling interval in the adaptive learning algorithm. Figure 7 is the output signal with SALA. To compare SALA with ALA, (6) and (11) are adopted, respectively, but all with the initial learning rate $\partial = 0.016$ and negative constant β. Figure 8 depicts the MSE of ALA and SALA.

Figures 7 and 8 show that the MSE of ALA and SALA are similar. The high MSE in Figure 8 before the 500th iterative time leads to inaccurate estimated curve in Figure 7

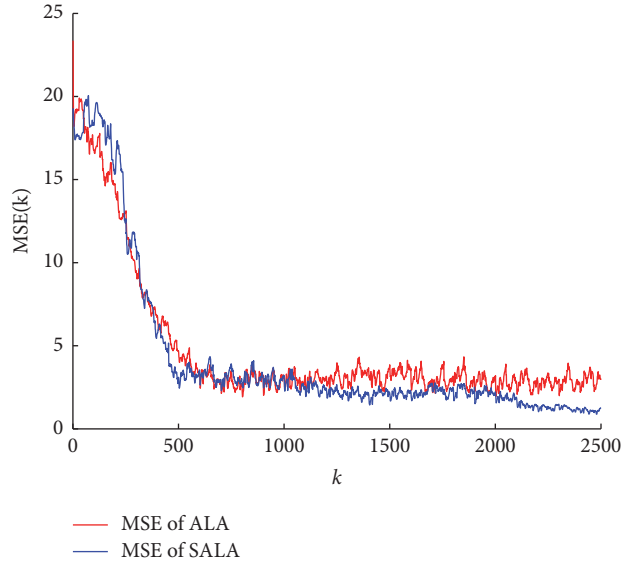


Figure 8: The MSEs of different algorithms.

Table 2: Performance of ALA and SALA for mobile signals.

	Convergence time (samples)	Mean steady state error	Computation time
ALA	582	2.3	96.6 s
SALA	586	2.0	64.2 s

correspondingly while the output estimated curve in Figure 7 becomes accurate when the steady state error in Figure 7 gets small. The results were evaluated using the performance index which is listed in Table 2. The convergence time and the mean steady state error of ALA and SALA are on the same level considering the random factor, but the computation time of SALA is obviously less than ALA. We can conclude from the analysis that the proposed SALA has advantage over ALA in the computation time.

4.2. Case 2. In order to verify the effectiveness of the proposed SALA, two music sources from the real environment are tested through the simulation. These music sources were mixed by random Gaussian noise matrix A of full rank $m = 2$. The mixtures were separated using SALA with the sampled variable step size. Figure 9 confirms the accurate and fast estimate as observed in Case 1.

5. Conclusions

Based on the discussion of fixed step-size algorithm and adaptive step-size algorithm for the blind separation of sources, a sampling adaptive step-size algorithm has been proposed. The algorithm has similar MSEs with adaptive step-size algorithm, but less computational time. By a smooth connection between two optimal points, the sampling method also has smooth curve and does not bring more recursion.

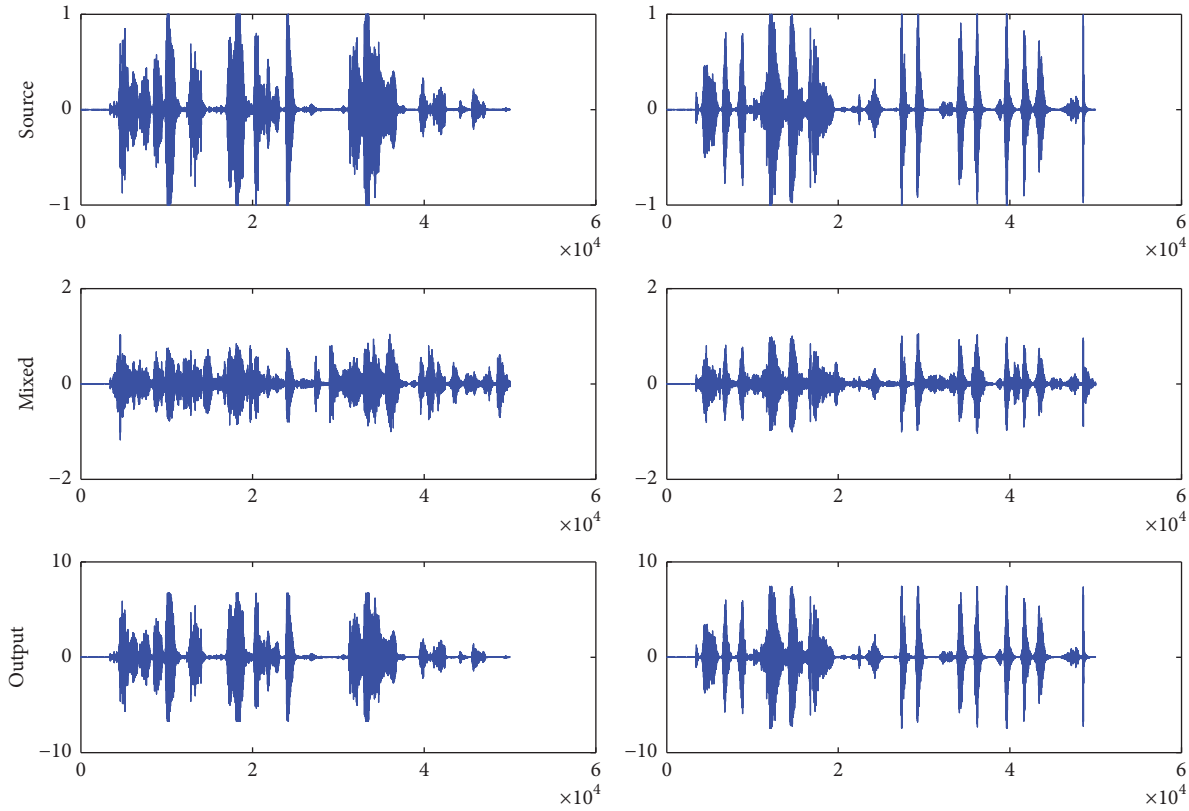


Figure 9: SALA for the real signal.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the Natural Science Foundation of China through the Grant 11702016.

References

[1] L. Ogiela, R. Tadeusiewicz, and M. R. Ogiela, "Cognitive computing in analysis of 2D/3D medical images," in *Proceedings of the 2007 International Conference on Intelligent Pervasive Computing, IPC 2007*, pp. 15–18, October 2007.

[2] Y. Zhang, M. Chen, N. Guizani, D. Wu, and V. C. Leung, "SOVCAN: safety-oriented vehicular controller area network," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 94–99, 2017.

[3] M. Chen, Y. Zhang, L. Hu, T. Taleb, and Z. Sheng, "Cloud-based wireless network: virtualized, reconfigurable, smart wireless network to enable 5G technologies," *Mobile Networks and Applications*, vol. 20, no. 6, pp. 704–712, 2015.

[4] Y. Zhang, "GroRec: a group-centric intelligent recommender system integrating social, mobile and big data technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[5] D. N. Levin, "Model-independent method of nonlinear blind source separation," in *Latent Variable Analysis and Signal Separation. LVA/ICA, 2017., Lecture Notes in Computer Science*, P. Tichavsk, M. Babaie-Zadeh, O. Michel, N. Thirion-Moreau, and P. Tichavský, Eds., vol. 10169, pp. 310–319, Springer, 2017.

[6] C. Hu, Q. Yang, M. Huang, and W. Yan, "Sparse component analysis-based under-determined blind source separation for bearing fault feature extraction in wind turbine gearbox," *IET Renewable Power Generation*, vol. 11, no. 3, pp. 330–337, 2017.

[7] N. D. Stein, "Nonnegative tensor factorization for directional blind audio source separation, Computer Science".

[8] E. Visser and L. Tewon, "Blind source separation in mobile environment using a priori knowledge," in *Proceedings of IEEE International Conference on Acoustics*, vol. 3, pp. 893–896, Quebec, Canada, 2004.

[9] S. M. Naqvi, Y. Zhang, and J. A. Chambers, "Multimodal blind source separation for moving sources," in *Proceedings of the ICASSP 2009 - 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 125–128, Taipei, Taiwan, April 2009.

[10] C. Ji, B. Tang, and J. Wang, "Blind source separation based on variable step length natural gradient algorithm," in *Proceedings of the 2011 3rd International Conference on Awareness Science and Technology, iCAST 2011*, pp. 261–264, September 2011.

[11] Y. L. Niu, J. C. Ma, and Y. Wang, "Blind separation for blurred images based on the adaptive nonholonomic natural gradient algorithm," in *Proceedings of the 2008 2nd International Symposium on Systems and Control in Aerospace and Astronautics, ISSCAA 2008*, December 2008.

[12] C. Ji, K. Yang, Y.-R. Wang, and M.-D. Liu, "Variable step-size nonholonomic natural gradient algorithm based on sign operator," *Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence*, vol. 27, no. 11, pp. 1026–1031, 2014.

[13] M. R. Bastian, J. H. Gunther, and T. K. Moon, "A simplified natural gradient learningalgorithm," *Advances in Artificial Neural Systems*, vol. 2011, pp. 3–13, 2011.

[14] J. Nan, "Natural gradient reinforcement learning algorithm with TD (λ)," *Computer Science*, vol. 37, no. 12, pp. 186–189, 2010.

[15] T. P. von Hoff and A. G. Lindgren, "Adaptive step-size control in blind souce separation," *Neurocomputing*, vol. 49, pp. 119–138, 2002.

[16] H. H. Dam, D. Rimantho, and S. Nordholm, "Second-order blind signal separation with optimal step size," *Speech Communication*, vol. 55, no. 4, pp. 535–543, 2013.

[17] S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication*, Springer, Boston, MA, USA, 2000.

[18] M. G. Jafari, J. A. Chambers, and D. P. Mandic, "A novel adaptive learning rate sequential blind source separation algorithm," *Signal Processing*, vol. 84, no. 4, pp. 801–804, 2004.

[19] J. Ce, Y. Peng, and Y. Yang, "Blind source separation based on improved natural gradient algorithm," in *Proceedings of the 2010 8th World Congress on Intelligent Control and Automation, WCICA 2010*, pp. 6804–6807, Jinan, China, July 2010.

[20] Y. Zhang, S. Lou, W. Zhang, and H. Chang, "Blind source separation algorithm of natural gradient based on estimation of score function," *Shuju Caiji Yu Chuli/Journal of Data Acquisition and Processing*, vol. 26, no. 2, pp. 167–171, 2011.

Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6210408, 9 pages
https://doi.org/10.1155/2018/6210408

Research Article

RADB: Random Access with Differentiated Barring for Latency-Constrained Applications in NB-IoT Network

Yiming Miao,[1] **Yuanwen Tian**,[1] **Jingjing Cheng**,[2,3] **M. Shamim Hossain**,[4] **and Ahmed Ghoneim**[4,5]

[1] *School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China*
[2] *School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China*
[3] *Graduate School of System Informatics, Kobe University, Kobe 657-8501, Japan*
[4] *Department of Software Engineering (SWE), College of Computer and Information Sciences (CCIS), King Saud University, Riyadh 11543, Saudi Arabia*
[5] *Computer Science Department, College of Science, Menoufia University, Menoufia 32721, Egypt*

Correspondence should be addressed to Jingjing Cheng; chengjj@hust.edu.cn

Received 10 September 2017; Revised 12 November 2017; Accepted 27 November 2017; Published 10 January 2018

Academic Editor: Huimin Lu

With the development of LPWA (Low Power Wide Area) technology, the emerging NB-IoT (Narrowband Internet of Things) technology is becoming popular with wide area and low-data-rate services. In order to achieve objectives such as huge amount of connection and wide area coverage within NB-IoT, the problem of network congestion generated by random access of numerous devices should be solved. In this paper, we first introduce the background of NB-IoT and investigate the research on random access optimization algorithm. Then we summarize relevant features of NB-IoT uplink and narrowband physical random access channel and design random access with differentiated barring (RADB), which can improve the insufficiency of traditional dynamic access class barring method. At last, the algorithms proposed in this paper are realized with established NB-IoT model using OPNET Modeler platform, and simulations are conducted. The simulation results show that RADB is able to effectively solve preamble request conflict generated by random access of numerous devices and preferentially provide efficient and reliable random access for latency-sensitive devices.

1. Introduction

With the increase of low-data-rate and low power services [1, 2], research on LPWAN communication technology develops correspondingly [3]. Based on whether the spectrum is licensed, LPWAN communication technologies [4] are divided into the following types: type 1 includes technologies that run in unlicensed frequency band, such as Lora and Sigfox. Mostly, these technologies are nonstandard and customized, so that safety cannot be guaranteed. Type 2 includes technologies that run in licensed frequency band, including mature technologies such as 2G/3G cellular communication technologies (GSM, CDMA, and WCDMA) and developing ones such as LTE and its evolution technology that are gradually deployed and applied at present which support terminals in various categories [5]. Basically, standards of these technologies are defined by international organizations for standard such as 3GPP (mainly formulates standards related to GSM, WCDMA, LTE, and its evolution technology) and 3GPP2 (mainly formulates standards related to CDMA).

NB-IoT is a kind of mass LPWA (Low Power Wide Area) technology put forth by 3GPP for application scenes with objectives of sensing and data acquisition (such as smart electric meter and environment monitoring [6]), characterized by advantages such as huge amount of connections, ultralow power consumption, wide area coverage, and mutual triggering between signaling [7, 8] and data [9–11]. In the meantime, it has support of excellent communication networks [12], such as cognitive vehicular networks [13] and cooperative communication networks [14].

In NB-IoT network, if a user equipment has access to the base station and sends service request, preamble request

transmitted in NPRACH (narrowband physical random access channel) should be considered first, that is, random access procedure. However, when many users request the same preamble, preamble conflict occurs. What is worse, if there are too many preamble conflicts in the network where huge amounts of users request NPRACH resources, network congestion would be caused inevitably. At that moment, huge amount of failures in random access and long-term latency in network would take place. Therefore, an optimized model for random access is extremely urgent in order to improve QoS of network and QoE of user [15].

In allusion to problem of network congestion caused by access of huge amount of devices [16] in M2M network [17], 3GPP determines the following alternatives: (1) access class barring schemes; (2) separate RACH resources for MTC; (3) dynamic allocation of RACH resources; (4) MTC specific backoff scheme; (5) slotted access; (6) pull based scheme. However, solutions mentioned above have not taken the aspect of latency in random access into consideration and thus cannot provide efficient and reliable random access for latency-sensitive devices, like application scene in [18–20]. Therefore, RADB is put forth in this paper to solve problems mentioned above.

The remainder of this paper is organized as follows. Section 2 provides some related research work. Section 3 introduces the NPRACH (narrowband physical random access channel) features of NB-IoT, including random access concept and traditional dynamic access class barring method. Section 4 shows the proposed RADB and envisioned NB-IoT architecture. Section 5 provides simulation setup and discusses experimental results. Finally, Section 6 concludes this paper.

2. Related Work

3GPP explicitly points out that it is necessary to give the theoretical computing model for uplink access latency when NB-IoT undertakes periodical and sudden MAR services [21]. Uplink access latency is composed of latencies in system synchronization, broadcast information reading, random access, resource allocation, data transmission and feedback response, and so on [7]. Among these latencies, some are deterministic processing latency, some are latencies related to signal detection, and others are random access latencies related to business activity [22]. Most projects of current research focus on computing of mean value and variance for random access latency; there are few projects of research that focus on probability density function (PDF) [23–25] of random access latency [26–28]. With quantity of waiting users and channel busy/idle as state variables, the moment generating function (MGF) for PDF of random access latency is deduced based on Markov process in [25, 29, 30]. But the problem of high computation complexity remains; it is even unsolvable when the quantity of users is too large. Reference [31] deduced PDF for random access latency on premise that time between arrivals and backoff obeys negative exponential distribution. References [23, 32] deduced PDF for random access latency on premise that retransmission times are fixed value or they obey geometric distribution. In research projects mentioned above, uniform distribution, exponential distribution, geometric distribution, and backoff mechanism are involved, but limitation of maximum retransmission times is taken into consideration only in [26, 33], which does not comply with actual protocol. The assumption that business models follow homogeneous Poisson or Bernoulli process is difficult to extend to application scenes of NB-IoT. In combination with 3GPP beta type business model, approximate form of PDF for random access latency, is given in [34] by estimating the maximum retransmission times for terminals with successful access through mean value of latency. In [35], the lower bound for random access latency is deduced through approximate beta distribution of piecewise linear function, but the effect of maximum retransmission times is not taken into consideration. In short, the theoretical computing model for random access latency has not been completely solved up to now, as well as the simplest Poisson business model and uniform backoff mechanism. Therefore, the research of statistic characteristics for random access latency in NB-IoT in any random access strength (two scenes C restricted PDCCH and unrestricted PDCCH C are taken into consideration, resp.) grows wide attention; not only should mean value and variance be included, but also its PDF and corresponding MGF should be deduced to perfect random access latency analysis theory for NB-IoT.

3. Narrowband Physical Random Access Channel Features

The transmission bandwidth for uplink of NB-IoT system is 180 kHz, supporting two kinds of subcarrier spacing: 3.75 kHz and 15 kHz. As for scenes with enhanced coverage, the subcarrier spacing of 3.75 kHz may provide larger system capacity when compared with subcarrier spacing of 15 kHz. However, in the scenes with internal operation mode, subcarrier spacing of 15 kHz has better compatibility with LTE, compared with subcarrier spacing of 3.75 kHz.

NB-IoT also reduces types of uplink physical channel, and some uplink physical channels are redesigned. Specifically, NB-IoT redesigns NPRACH and does not support PUCCH (Physical Uplink Control Channel).

3.1. NPRACH Features. Because the bandwidth of traditional LTE physical random access channel (PRACH) is 1.08 MHz, which exceeds restriction on bandwidth of 180 kHz for uplink of NB-IoT, random access channel is redesigned and named as NPRACH. An NPRACH preamble is composed of four symbol groups, and each symbol group is composed of one cyclic prefix (CP) and five symbols. The CP with length of 66.67 s (Format 0) is suitable for cell with radius of 10 km, and the CP with length of 266. 7 s (Format 1) is suitable for cell with radius of 40 km; thus the objective of coverage gain is achieved. The value of each symbol is fixed to 1, and modulation is conducted at subcarrier spacing of 3.75 kHz (with symbol duration of 266.67 s). Thereinto, the frequency modulation index for each symbol group is different. However, the waveform of NPRACH preamble follows single-tone frequency hopping. Figure 1 shows a case of NPRACH frequency hopping [36]. In order to support

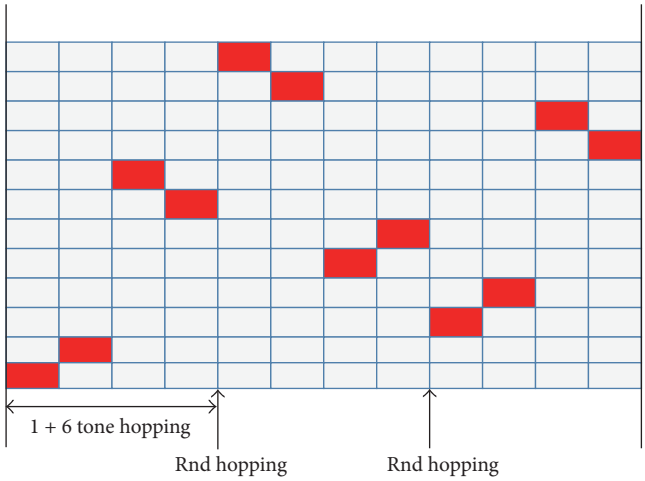


Figure 1: NPRACH frequency hopping.

coverage gain, the repeated use of one NPRACH preamble will be permitted for as many as 128 times.

3.2. Random Access. In allusion to requirements of coverage enhancement, random access based on coverage level is adopted by NB-IoT system. The terminal will judge the current coverage level based on signal strength measured, and appropriate random access resource shall be chosen based on corresponding coverage level to launch random access. In order to meet requirements on data transmission under different coverage levels, different times of repetition, transmission cycles, and so on will be allocated to each coverage level by base station. For example, the terminal under poor coverage level needs to adopt more times of repetition to guarantee correct transmission of data, while large transmission cycle may be allocated in order to prevent terminals under poor coverage level from occupying too many system resources.

Effective access control mechanism is required to guarantee and control access of terminal and preferential access of some abnormal data due to the huge amount of IoT terminals. As for access control mechanism of NB-IoT system, the EAB mechanism of LTE system (SIB14) and backoff mechanism of random access procedure are used for reference. Also, MIB-NB broadcasts indication of access control to reduce power consumption of SIB14-NB that tries to read at terminal.

In NB-IoT, random access is used in many aspects such as initial access during establishing wireless link [37] and scheduling request [38]. A main objective of random access is to realize synchronization of uplink, which plays a vital role in maintaining orthogonality of uplink. Similar to random access mechanism of LTE, competition-based random access procedure of NB-IoT includes the following four steps: (1) user equipment (UE) sends a random access preamble; (2) a random access response (including timing advance command and uplink resources scheduling) will be transmitted by network for use of UE in the third step; (3) UE broadcasts its identity label in the network with scheduled resources; (4) contention-resolution message is transmitted by network to solve conflict caused when multiple UE pieces send the same random access preamble in the first step.

In order to better serve UE pieces under different coverage levels and with different degrees of path loss, up to 3 kinds of different NPRACH resources will be allocated in a cell by NB-IoT network. In each kind of allocation, each basic random access preamble has a given duplicate value for repeated use. UE will measure its signal receiving power at downlink to estimate its coverage level and adopt NPRACH resources allocated by network to send random access preamble for the estimated coverage level. In order to deploy NB-IoT network in different scenes, flexible allocation of NPRACH resources under time-frequency resource grid is allowed by NB-IoT; the specific parameters are as follows:

(i) Time domain: NPRACH resource has periodicity referring to the start time of NPRACH resource in a period of time.

(ii) Frequency domain: it includes frequency distribution (determined by subcarrier migration) and quantity of subcarrier.

In early field test and deployment of NB-IoT, some UE pieces do not support multitone transmission. Therefore, before transmission scheduling for uplink, the network should know multitone transmission capacity of UE. In addition, in the first step of random access procedure, a UE should express information on whether it supports multitone transmission, so that transmission scheduling for uplink can be realized by the network in the third stage of random access procedure. To be specific, network divides NPRACH subcarriers into two nonoverlapping sets by their frequency domain. In the third step of random access procedure, UE may choose one of the two sets to send its signal of random access preamble and thus to express whether it supports multitone transmission.

Consequently, UE determines its coverage level by measuring signal receiving power at downlink. After reading system information allocated by NPRACH resources, UE is able to conduct NPRACH resource allocation and to set retransmission times required by estimating its coverage level and transmission power of random access preamble. Then, UE is able to continuously and repeatedly transmit basic single-tone random access preamble within a period of NPRACH resources.

However, continuous retransmission of single-tone random access preamble in a single cycle may cause preamble request conflict. With a large amount of conflicts increasing the request and response delay time (i.e., random access latency would be longer), the network will fall into congestion inevitably. Access request of huge amount of devices will bring great challenge to wireless access capacity of access network while the main focus is on overload problem in a cell as for congestion in access network. For example, assuming that large amounts of devices access a cell simultaneously, the conflict probability of access channel in that cell will increase rapidly, and severe paralysis will be caused if control is not available in time.

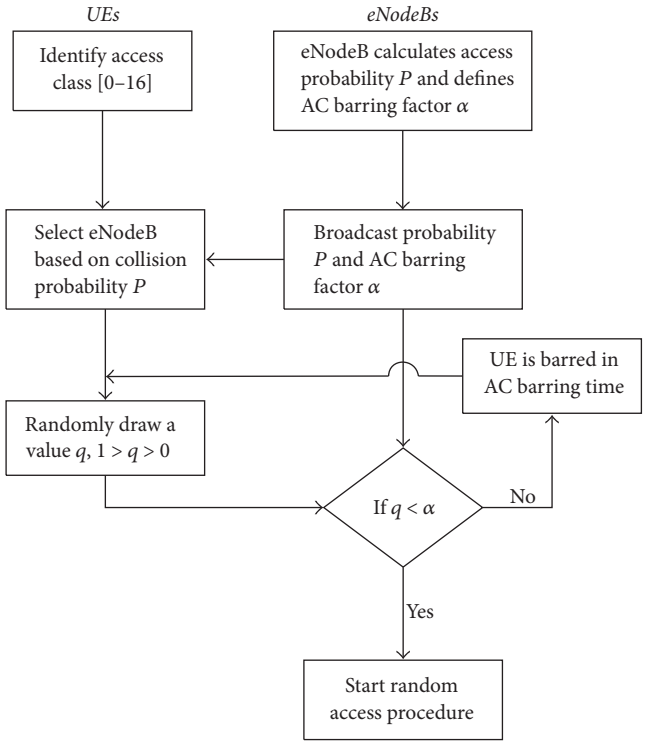


Figure 2: Access class barring schemes.

3.3. Access Class Barring Schemes. The solution to solve network congestion caused by large amount of access requests in cellular machine-to-machine (M2M) network is put forth in [39], that is, dynamic access class barring (ACB) method. Figure 2 shows the ACB method, including base station selection and load balance strategy. Main steps are as follows:

(i) Step 1: eNodeB evaluates the conflict probability P of random access based on the arrival rate and transmission rate of data package. Access class barring parameter α depends on PRACH condition (total preamble number and current requests number for preamble).

(ii) Step 2: every eNodeB periodically broadcasts conflict probability P and AC barring parameter α.

(iii) Step 3: UE chooses the corresponding eNB when the maximum access success probability is achieved and conducts connection and communication to the eNB.

(iv) Step 4: each UE will generate a random number q ($1 > q > 0$), and start/prohibit random access procedure following ACB mechanism; that is, UE is able to successfully start random access procedure when and only when $q < \alpha$.

4. Random Access with Differentiated Barring and Network Architecture

4.1. Random Access with Differentiated Barring. In RADB put forth in this paper, conflict probability, random number generation mechanism, AC barring parameter α, and access recognition algorithm of traditional ACB scheme are defined specifically; then corresponding improvement is made.

Devices in NB-IoT network are divided into Class A and Class B. Class A stands for latency-sensitive devices (that long for low data transmission latency), while Class B stands for non-latency-sensitive devices (that may tolerate longer data transmission latency). NPRACH period T is divided into t time slots and random number $q_{(t)}$ of UE pieces in each time slot is defined in formula (1), where $t = 0, 1, 2, \ldots, T$.

$$\begin{aligned} &q_{A(t)}, \quad t = 0, 1, 2, \ldots, T; \\ &q_{B(t)} \in (0, 1), \quad t = 0, 1, 2, \ldots, T. \end{aligned} \tag{1}$$

In allusion to preamble defined in eNodeB, we assume the sum is S, request number of current preamble is x, and channel conflict probability is P, as shown in the following formula:

$$P = 1 - \left(1 - \frac{1}{S}\right)^{(x-1)}. \tag{2}$$

If $P = 1 - (1 - 1/S)^{(x-1)} < 0.1$, it is held that the channel conflict probability is low; that is, success rate for random access is high. In this case, it is assumed that the maximum value of x is X, and X stands for the maximum request number of preamble when success rate for random access is guaranteed. Let number of Class A devices be N_A and number of Class B devices be N_B; then the total number of devices $N = N_A + N_B$, and the number of devices with successful access $N_s = N_A + N_B$. Therefore, dynamic AC barring parameter α may be set as per the following formula:

$$\alpha = \begin{cases} 1, & N < X; \\ \dfrac{X - N_A}{N_B}, & N > X > N_A; \\ 0, & N_A > X. \end{cases} \tag{3}$$

Because $q_{A(t)}$ is always equal to 0 (remaining constant while t changes), $q_{A(t)}$ is always less than or equal to α; in other words, devices in Class A have the right to start random access procedure preferentially. Only when redundancy in sum of preambles (sum of preambles is more than number of devices in Class A that request preamble) appears, could devices in Class B have the chance to access NB-IoT eNodeB. The pseudocode of RADB Scheme is shown as Algorithm 1.

4.2. NB-IoT Network Architecture. In accordance with previous research [40], the architecture of NB-IoT network established based on OPNET Modeler platform in this paper is as shown in Figure 3, mainly including 5 parts: NB-IoT terminal, NB-IoT access network, NB-IoT core network, NB-IoT cloud platform, and vertical industry center [41, 42]. In order to embody coverage gain attribute of NB-IoT network and three corresponding NPRACH resource configuration options, the whole network is divided into 3 areas from long range to short range, and 3 MCSs (MCS 9, MCS 20, and MCS28) are selected, respectively, as modulation and coding strategy for each area based on MCS index table given in [40]; MCS index ID is made as area ID. In each area, NB-IoT

Require: Class A: delay-sensitive device;
Class B: non-sensitive devices;
Ensure:
for $t = 0, t < T, ++t$ **do**
Delay-sensitive devices generate a parameter $q_{A(t)} == 0$;
Non-sensitive devices randomly generate a parameter $q_{B(t)}$;
if $N < X$ **then**
$\alpha = 1$;
ALL devices are not barred in current period and can randomly select preambles;
$N = N - N_t$;
else $\{N_A < X\}$
$\alpha = \frac{X - N_A}{N_B}$;
Delay-sensitive devices will not be barred in current period and can randomly select preambles;
if N_t is non-sensitive device & $q_{B(t)} < \alpha$ **then**
N_t can start random access procedure and select preambles;
end if
$N_S = N_A + \alpha N_{Bt}$;
$N_A = N_A - N_{At}$ or $N_B = N_B - N_{Bt}$;
else $\{N_A > X\}$
$\alpha = 0$;
Only delay-sensitive devices will not be barred in current period and can randomly select preambles.
Non-sensitive devices will be barred in current period.
end if
end for

Algorithm 1: Random access with differentiated barring.

terminals (UE pieces, standing for devices carried by different mobile users) in different numbers are deployed; local ID in an area is made as identity label of these devices in that area. All user devices in the network send random access request to corresponding NB-IoT base station (improved LTE base station, eNodeB); whether an equipment could successfully enter random access procedure is determined through RADB.

As shown in Figure 4, when the synchronous relationship between an NB-IoT terminal and base station is not established, that terminal must send random access request before it could access network, that is, from idle state to connected state. At that moment, it is difficult for limited system information and channel information to guarantee reliability of transmission with closed loop random access control; therefore, Algorithm 1 is adopted to improve and set up random access process model in physical layer in this paper.

5. Simulation and Analysis

Table 1 lists the values of important parameters considered in the simulations [40]. These values were selected to reflect real-world implementations of NB-IoT network and based on our previous research [40]. The simulations were run multiple times and the presented results are an average of these runs.

Firstly, comparison is made between access success rate of NB-IoT network with different quantities of equipment under RADB and that of NB-IoT network under traditional access class barring scheme when barring parameter α is 0.2 or 0.8, as shown in Figure 5. When the quantity of equipment is less than 150, the performance of ACB scheme with α of 0.2 is still fine. However, with increase in quantity of equipment, the RADB that dynamically adjusts barring parameter shows strong controlling force over random access; it effectively makes sure that latency-sensitive equipment could successfully access network to the largest degree.

Figure 6 shows the comparative results for access latency generated with different random access models in NB-IoT network with 350 mobile devices; the random access models include RADB and traditional access class barring scheme with barring parameter α equal to 0.2, 0.4, 0.6, and 0.8. It is shown that the latency for ACB scheme with barring parameter α of 0.8 is the longest (latency for ACB scheme with barring parameter α of 0.6 is the second longest), which indicates that large amounts of devices access network at that time leading to the increasing probability of channel conflict and frequent network congestion. However, though the latency for ACB scheme with barring parameter α of 0.2 and 0.4 is short, the quantity of devices that can successfully access network at that time is too small; therefore instability of network is caused. As for RADB, with setting of dynamic

Table 1: Simulation parameters.

Element	Attribute	Value
EPS	QoS class identifier	1 (GBR)
	Allocation retention priority	2
	Uplink guaranteed bit rate	32 Kbps
	Downlink guaranteed bit rate	96 Kbps
	Uplink maximum bit rate	32 Kbps
	Downlink maximum bit rate	384 Kbps
Physical layer profiles	UL SC-FDMA channel	
	Base frequency	1920 MHz
	Bandwidth	0.2/3/5/10/15/20 MHz
	Cyclic prefix type	7 symbols per slot
	DL OFDMA channel	
	Base frequency	2110 MHz
	Bandwidth	0.2/3/5/10/15/20 MHz
	Cyclic prefix type	7 symbols per slot
eNodeB	Failure/recovery specification time	200 seconds
	Barring parameter α of ACB	0.2/0.4/0.6/0.8
UE pieces	Battery capacity	5
	Maximum transmission power	10 mW
	Number	50,500 per 50
	Modulation and coding scheme index	9/20/28
	Operating power	100 mW

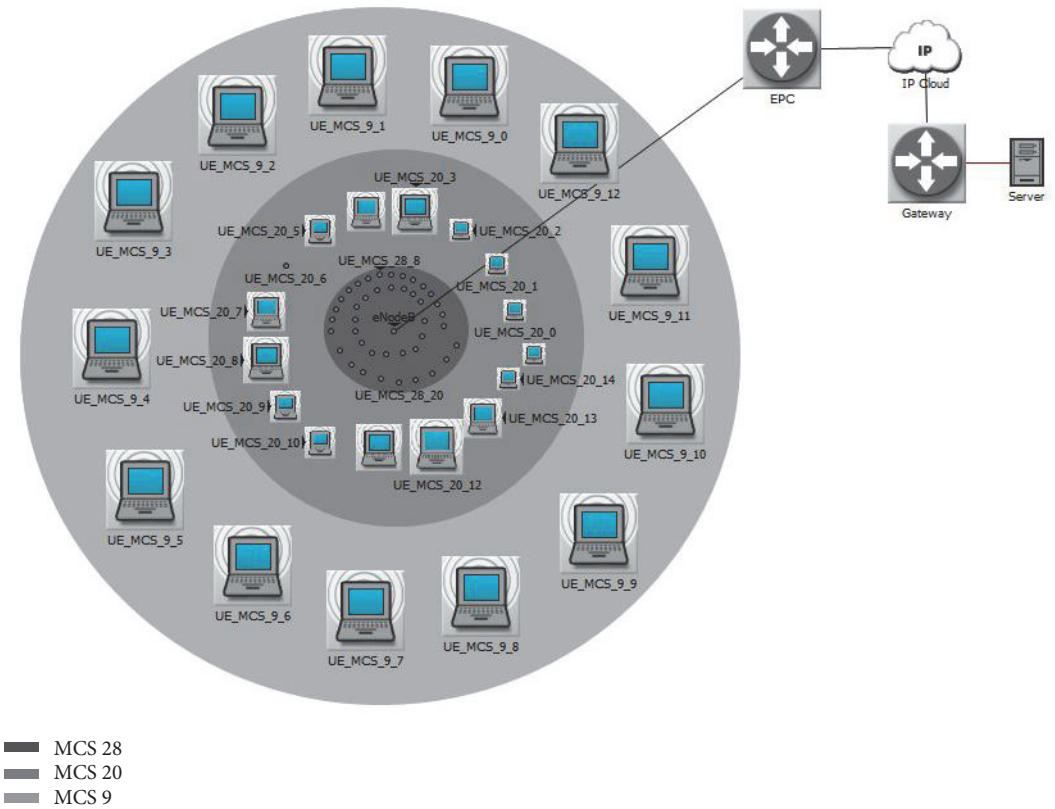


Figure 3: NB-IoT network architecture.

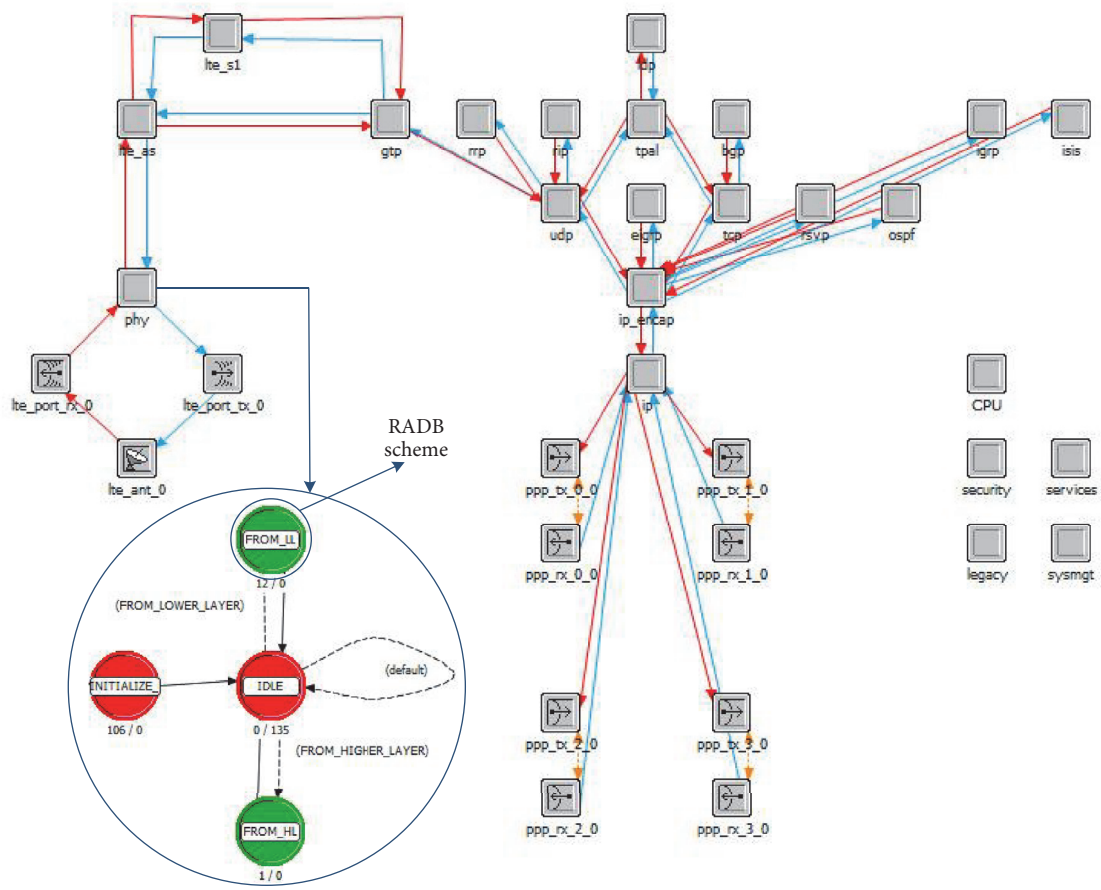


Figure 4: RADB process model.

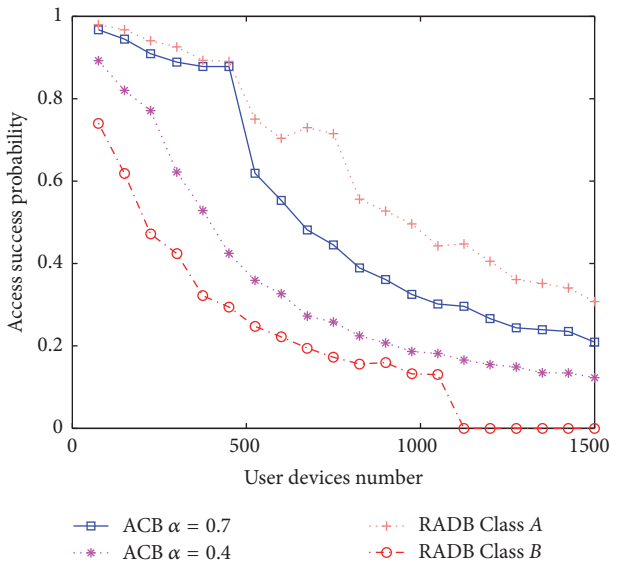


Figure 5: Access success probability.

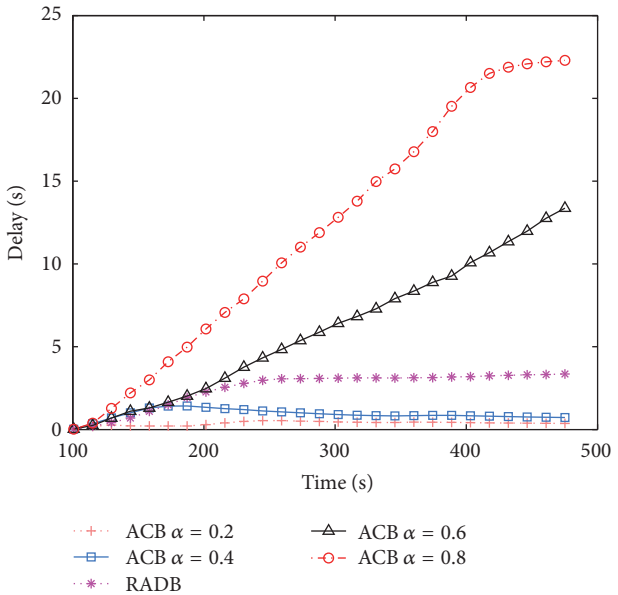


Figure 6: Access delay.

barring parameter, network could enter stable state earlier; thus network latency is controlled effectively.

Figure 7 shows the comparative results for network load generated with different random access models in NB-IoT

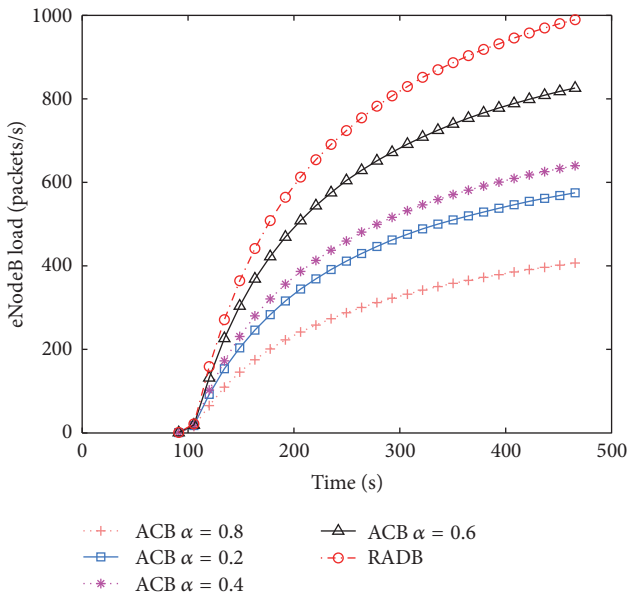


Figure 7: eNodeB load.

network with 350 mobile devices. The throughput (loading capacity) of a system is closely related to the consumption of CPU by request, peripheral interface, IO, and so on. If the consumption of CPU by single request is higher, the response rate of peripheral interface and IO is slower and the throughput of the system is lower. This situation is the opposite when the CPU consumption is very low. It can be seen that the response rate of base station is very slow when ACB scheme α is equal to 0.8 due to long access latency; therefore, the network throughput is very low. When ACB scheme α is equal to 0.2, though access latency is short, devices that access network are fewer, so the network throughput is not high. However, as for RADB put forth in this paper, because network latency is controlled, the requirements of latency-sensitive devices on network are met; thus network throughput is guaranteed.

6. Conclusion

In this paper, the background of NB-IoT is introduced and worldwide research related to optimized algorithms for random access is investigated. Then, characteristics related to NB-IoT uplink and narrowband physical random access channel are summarized, improvement is made in allusion to insufficiency of traditional dynamic access class barring method, and RADB is designed. Furthermore, the algorithms put forth in this paper are realized with established NB-IoT model, and simulation experiment is conducted. The results of simulation experiment show that RADB is able to effectively solve preamble request conflict generated by random access of numerous devices and to preferentially provide efficient and reliable random access for latency-sensitive devices.

Nevertheless, problems of channel resource distribution and resource utilization rate are not taken into consideration in algorithm put forth in this paper. In subsequent research, we will continue to study equipment access algorithms with high energy efficiency and low load and improve existing models, thus providing theoretical and experimental bases for future large-scale deployment of NB-IoT network in the real world. Also, how to use Big Data techniques [43] to support the NB-IoT based services is still interesting but challenging.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia, for funding this work through the research group Project no. RGP-229.

References

[1] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular fog computing: a viewpoint of vehicles as the infrastructures," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3860–3873, 2016.

[2] F. Xu, Y. Li, H. Wang, P. Zhang, and D. Jin, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 1147–1161, 2017.

[3] M. Chen, Y. Miao, Y. Hao, and K. Hwang, "Narrow Band Internet of Things," *IEEE Access*, vol. 5, pp. 20557–20577, 2017.

[4] X. Xiong, K. Zheng, R. Xu, W. Xiang, and P. Chatzimisios, "Low power wide area machine-to-machine networks: Key techniques and prototype," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 64–71, 2015.

[5] M. Chen, Y. Hao, L. Hu, K. Huang, and V. Lau, "Green and Mobility-aware Caching in 5G Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8347–8361, 2017.

[6] K. Lin, M. Chen, J. Deng, M. Hassan, and G. Fortino, "Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 3, pp. 1294–1307, 2016.

[7] 3GPP TR 45.820, "Cellular system support for ultra-low complexity and low throughput cellular internet of things," 2015.

[8] 3GPP TS 36.211, "E-UTRA Physical channels and modulation-Chap.10 Narrowband IoT," 2016.

[9] C.-C. Tseng, H.-C. Wang, F.-C. Kuo, K.-C. Ting, H.-H. Chen, and G.-Y. Chen, "Delay and power consumption in LTE/LTE-A DRX mechanism with mixed short and long cycles," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1721–1734, 2015.

[10] R. Cheng, A. Deng, and F. Meng, *Study of NB-IoT Planning Objectives And Planning Roles*, China Mobile Group Design Institute Co., 2016.

[11] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the vehicular networks," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1932–1940, 2017.

[12] K. He, J. Chen, R. Du, Q. Wu, G. Xue, and X. Zhang, "DeyPoS: Deduplicatable Dynamic Proof of Storage for Multi-User

Environments," *IEEE Transactions on Computers*, vol. 65, no. 12, pp. 3631–3645, 2016.

[13] D. Tian, J. Zhou, Z. Sheng, and V. C. M. Leung, "Robust Energy-Efficient MIMO Transmission for Cognitive Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3845–3859, 2016.

[14] D. Tian, J. Zhou, Z. Sheng, M. Chen, Q. Ni, and V. C. Leung, "Self-Organized Relay Selection for Cooperative Transmission in Vehicular Ad-Hoc Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9534–9549, 2017.

[15] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, "Trajectory Recovery From Ash: User Privacy Is NOT Preserved in Aggregated Mobility Data," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 1241–1250, 2017.

[16] J. Chen, K. He, Q. Yuan, G. Xue, R. Du, and L. Wang, "Batch identification game model for invalid signatures in wireless mobile networks," *IEEE Transactions on Mobile Computing*, vol. 16, no. 6, pp. 1530–1543, 2017.

[17] M. Chen, J. Wan, S. Gonzalez, X. Liao, and V. C. M. Leung, "A survey of recent developments in home M2M networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 98–114, 2014.

[18] M. Chen, Y. Hao, M. Qiu, J. Song, D. Wu, and I. Humar, "Mobility-aware caching and computation offloading in 5G ultra-dense cellular networks," *Sensors*, vol. 16, no. 7, pp. 974–987, 2016.

[19] M. Chen, J. Yang, X. Zhu, X. Wang, M. Liu, and J. Song, "Smart home 2.0: innovative smart home system powered by botanical IoT and emotion detection," *Mobile Networks and Applications*, vol. 22, pp. 1159–1169, 2017.

[20] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[21] 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.

[22] K. Lin, J. Song, J. Luo, W. Ji, M. Shamim Hossain, and A. Ghoneim, "Green Video Transmission in the Mobile Cloud Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 1, pp. 159–169, 2017.

[23] Y. Yang and T.-S. P. Yum, "Delay Distributions of Slotted ALOHA and CSMA," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1846–1857, 2003.

[24] C.-H. Wei, P.-C. Lin, and R.-G. Cheng, "Comment on 'an efficient random access scheme for OFDMA systems with implicit message transmission'," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 414-415, 2013.

[25] A. Mutairi, S. Roy, and G. Hwang, "Delay analysis of OFDMA-aloha," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 89–99, 2013.

[26] F. A. Tobagi, "Distributions of packet delay and interdeparture time in slotted ALOHA and Carrier Sense Multiple Access," *Journal of the ACM*, vol. 29, no. 4, pp. 907–927, 1982.

[27] J.-B. Seo and V. C. M. Leung, "Design and analysis of backoff algorithms for random access channels in UMTS-LTE and IEEE 802.16 systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3975–3989, 2011.

[28] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5G," *IEEE Network*, vol. 30, no. 1, pp. 44–51, 2016.

[29] L. Dai, "Stability and delay analysis of buffered aloha networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2707–2719, 2012.

[30] K. Zheng, F. Liu, L. Lei, C. Lin, and Y. Jiang, "Stochastic performance analysis of a wireless finite-state Markov channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 782–793, 2013.

[31] M. E. Rivero-Angeles, D. Lara-Rodríguez, and F. A. Cruz-Pérez, "Access delay analysis of adaptive traffic load - Type protocols for S-ALOHA and CSMA in EDGE," in *Proceedings of the 2003 IEEE Wireless Communications and Networking Conference: The Dawn of Pervasive Communication, WCNC 2003*, vol. 3, pp. 1722–1727, March 2003.

[32] M. E. Rivero-Angeles, D. Lara-Rodriguez, and F. A. Cruz-Perez, "Gaussian approximations for the probability mass function of the access delay for different backoff policies in S-ALOHA," *IEEE Communications Letters*, vol. 10, no. 10, pp. 731–733, 2006.

[33] R. R. Tyagi, F. Aurzada, K.-D. Lee, and M. Reisslein, "Connection Establishment in LTE-A networks: Justification of poisson process modeling," *IEEE Systems Journal*, vol. 99, pp. 1–12, 2015.

[34] C.-H. Wei, G. Bianchi, and R.-G. Cheng, "Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 4, pp. 1940–1953, 2015.

[35] M. Koseoglu, "Lower Bounds on the LTE-A Average Random Access Delay under Massive M2M Arrivals," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2104–2115, 2016.

[36] Y. E. Wang, X. Lin, A. Adhikary et al., "A Primer on 3GPP Narrowband Internet of Things," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 117–123, 2017.

[37] S.-S. Kim, S. McLoone, J.-H. Byeon, S. Lee, and H. Liu, "Cognitively Inspired Artificial Bee Colony Clustering for Cognitive Wireless Sensor Networks," *Cognitive Computation*, vol. 9, no. 2, pp. 207–224, 2017.

[38] H. Liu, A. Abraham, V. Snášel, and S. McLoone, "Swarm scheduling approaches for work-flow applications with security constraints in distributed data-intensive computing environments," *Information Sciences*, vol. 192, pp. 228–243, 2012.

[39] L. Ferdouse and A. Anpalagan, "A dynamic access class barring scheme to balance massive access requests among base stations over the cellular M2M networks," in *Proceedings of the 26th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2015*, pp. 1283–1288, September 2015.

[40] Y. Miao, W. Li, D. Tian, M. S. Hossain, and M. F. Alhamid, "Narrow Band Internet of Things: Simulation and Modelling," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1-1, 2017.

[41] Y. Li and M. Chen, "Software-defined network function virtualization: a survey," *IEEE Access*, vol. 3, pp. 2542–2553, 2015.

[42] Y. Li, F. Zheng, M. Chen, and D. Jin, "A unified control and optimization framework for dynamical service chaining in software-defined NFV system," *IEEE Wireless Communications Magazine*, vol. 22, no. 6, pp. 15–23, 2015.

[43] X. Wang, Y. Zhang, V. C. M. Leung, N. Guizani, and T. Jiang, "D2D big data: content deliveries over wireless device-to-device sharing in realistic large scale mobile networks," *IEEE Wireless Commun*, vol. 25, no. 1, pp. 1–10, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 2306458, 6 pages
https://doi.org/10.1155/2017/2306458

Review Article

A Systematic Review of Security Mechanisms for Big Data in Health and New Alternatives for Hospitals

Sofiane Hamrioui,[1] Isabel de la Torre Díez,[2] Begonya Garcia-Zapirain,[3] Kashif Saleem,[4] and Joel J. P. C. Rodrigues[5,6,7,8]

[1]*Bretagne Loire and Nantes Universities, UMR 6164, IETR Polytech Nantes, Nantes, France*
[2]*Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain*
[3]*University of Deusto, Avenida de las Universidades 24, 48007 Bilbao, Spain*
[4]*Center of Excellence in Information Assurance (CoEIA), King Saud University, Riyadh, Saudi Arabia*
[5]*National Institute of Telecommunications (Inatel), Santa Rita do Sapucaí, MG, Brazil*
[6]*Instituto de Telecomunicações, Lisboa, Portugal*
[7]*University of Fortaleza (UNIFOR), Fortaleza, CE, Brazil*
[8]*University ITMO, St. Petersburg, Russia*

Correspondence should be addressed to Isabel de la Torre Díez; isator@tel.uva.es

Received 21 September 2017; Accepted 30 October 2017; Published 4 December 2017

Academic Editor: Yin Zhang

Computer security is something that brings to mind the greatest developers and companies who wish to protect their data. Major steps forward are being taken via advances made in the security of technology. The main purpose of this paper is to provide a view of different mechanisms and algorithms used to ensure big data security and to theoretically put forward an improvement in the health-based environment using a proposed model as reference. A search was conducted for information from scientific databases as Google Scholar, IEEE Xplore, Science Direct, Web of Science, and Scopus to find information related to security in big data. The search criteria used were *"big data", "health", "cloud",* and *"security"*, with dates being confined to the period from 2008 to the present time. After analyzing the different solutions, two security alternatives are proposed combining different techniques analyzed in the state of the art, with a view to providing existing information on the big data over cloud with maximum security in different hospitals located in the province of Valladolid, Spain. New mechanisms and algorithms help to create a more secure environment, although it is necessary to continue developing new and better ones to make things increasingly difficult for cybercriminals.

1. Introduction

Big data refers to the huge amount of information that is created and proves difficult to analyze in real time, to the extent that traditional databases are not sufficient for dealing with such data [1]. When big data is mentioned, it tends to refer to the three Vs (volume, velocity, and variety) [2], and some even extend this to the five Vs: volume, velocity, variety, veracity, and value [3, 4]. These five characteristics mentioned are defined as follows: volume refers to the size of the data generated, although certain minimums are yet to be established as this is a relative concept. Velocity refers to a large amount of data generated over time. Variety refers to a combination of different information formats, whether structured, semistructured, or without structure. Veracity refers to the fact that none of the data generated is of any use if it is not reliable. Lastly, value refers to the scientific value attributed to this data [5–7].

The term "big data" has been exponentially growing in use since 2011 and is taking on increasing weight in both society and the world of business [8]. As this is a new concept, the lack of trust factor is involved that raises other associated problems [9]. Some of these problems that are seen in the use of big data are rather dangerous. Even though the data is

stored anonymously, there is little control over it, whether this data is private or personal [10]. Moreover, there are benefits in using big data, the most important of which is that it enables the Government to improve quality of life in society via the analysis of vast amounts of information [10].

Traditional databases use standard SQL that generates request and handle relational tables. Owing to the relationships that exist between the tables, utilizing them is not practical because the big data mainly refers to the unstructured information [10]. NoSQL databases have emerged as a result and offer better functionality especially for the purpose of storing and maintaining information on a large scale without being concerned with the format of data in which it is presented. Furthermore, the NoSQL offers high performance for the large volumes of heterogeneous data within a distributed environment [11–15].

Moreover, the terms big data and cloud computing have enormously gained prominence in recent years, and one of the reason for this is because big data is directly related to the cloud [12, 13]. The importance is because the cloud is coming up with the new architecture paradigms in information technology [16–18]. Emerging cloud computing technology offers a solution to reduce the cost of development and operating mobile networks [19–22].

As we have mentioned above, the big data uses cloud and is accessible all around the world from any object with the help of Internet [14]. This raises a big question of security; therefore below a summary of the different levels of big data security over cloud is given. This paper extensively reviews the work carried out by different experts to minimise the risks in utilizing cloud in handling big data [23–26]. Secondly, the concrete solutions are presented to improve the security of big data in a healthcare scenario.

The next section covers the methodology used to obtain and filter the related information required. In Section 3.1, the recent security mechanisms for healthcare big data in cloud are reviewed. Additionally, in Section 3.2 the security model is given and is explained in detail. Lastly, Section 4 provides the conclusion.

2. Methods

An exhaustive search was conducted in order to carry out the research of papers on some of the most important and commonly used websites and scientific databases, namely, Google Scholar, IEEE Xplore [27], Scopus [28], Science Direct [29], and Web of Science [30].

A specific search was conducted in each of them using the following words: "cloud" AND/OR "big data" AND "health" AND "security" in the title and abstract. In all cases, the time span was from 2008 to September 2017. Figure 1 shows the 3169 results obtained via the searches conducted, and we also refer to any papers that were disregarded as a result of their being duplicates or having a title that is unrelated to our area of interest. We ended up with 22 papers after having read 134 and seeing which of them proved beneficial to us after viewing the corresponding abstracts.

All the papers related to health. At the end, we decided to use these papers because after reading the others, although initially appearing to cover big data and associated security, we noted that they only mention the big data situation or the situation regarding legislation in Europe and the USA governing data on the cloud.

Therefore, we decided to disregard these as they failed to provide us with relevant information, and in making our selection, we took papers written in English into consideration. After reading the titles and abstract and ending up with the 134 papers mentioned, we then proceeded to read their content and thus determine which of them would provide us with information related to big data and its security on the cloud or on databases it uses to store data.

3. Results and Discussion

3.1. Security Mechanisms Based Literature Review. In this section, we provide a summary of the main advances made by the scientific community that will help to keep healthcare data more secure. Recently, Wang et al. (2017) enhance attribute-based encryption (ABE) that is introduced by the Cloud Security Alliance (CSA). The improved auxiliary input model based Ciphertext-policy ABE (CP-ABE) and key-policy ABE (KP-ABE) schemes are presented. While conducting the comparison, the improved model considers also the encryptor leakage (leakage of randomness) in front of other auxiliary input model. Furthermore, an improved strong extractor from the modified Goldreich–Levin theorem is given. The performance comparison is conducted between the other three leakage resilient CP-ABE and the proposed schemes. The author has programmed the CP-ABE scheme in C language and has implemented it on two different types of processors based platforms by using the pairing based cryptography (PBC) library to test the encryption time.

Cho et al. (2016) provide us with architecture based on a double layer for the working environment with big data. These two layers are the prefiltering layer and the postfiltering layer. The first-mentioned is in charge of searching for and eliminating sensitive personal information from the data gathered, which is done in order to make the information anonymous and thus make it more difficult to identify the person in particular. The second, postfiltering layer, disguises the summarized sensitive information following big data analysis [31].

Liu et al. (2015) define a series of steps in which the validity of the data stored is verified externally. External verification is as important as the security provided by the server and, as this is an external agent, certain steps need to be established to maintain data security [32].

Fabiano et al. (2015), from the University of Wyoming, have been developing a variant of the MapReduce paradigm to be applied in security which complies with HIPAA, as well as using the OpenSSL encryption package. To ensure maximum scalability, they implemented a hybrid of the OpenMP-MPL programming paradigm, by means of which they enabled each processing core to be assigned a number of files and then for each core to subdivide these files, depending on the number of threads being used [33].

Yan et al. (2016) propose two security schemes in which the aim is to protect the confidential information of

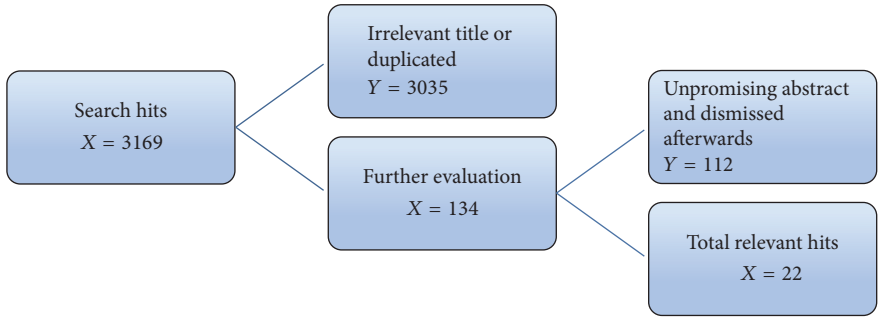


Figure 1: Flow chart of the steps followed in the review.

trusted suppliers. The first scheme focuses on computational efficiency while the second provides better protection at the expense of computational cost. They use proxy-based additive homomorphism with reencryption to design these two schemes for Privacy-Preserving Trust Evaluation (PPTE) [34].

Zhou et al. (2015) propose an encryption algorithm that focuses on image security. This algorithm is based on the Chaos stochastic process and the Line Map principle and is designed to ensure that if an image is encrypted, nothing will continue to be seen if an attempt is made to decrypt the key. The disadvantage of this algorithm is that it can only be used for those images that have the same width and length, although the authors have carried out tests and reached the conclusion that it is a robust algorithm [35].

Hsu et al. (2014) show us how to develop a protocol for the secure transfer of data and also propose a protocol for the transfer of a group key. To do so, they create a variant of the Diffie Hellman algorithm which is designed for one-to-one communication rather than between several individuals. The motivation behind this protocol is to preserve the key refresh, its confidentiality, and its authentication [36].

Jing (2014) comments on the growing use of the cloud as a storage space. To improve security, they propose a double encryption data system, which consists of an initial encryption using the AES encryption algorithm and therefore a symmetric algorithm. They then use the RSA algorithm as an asymmetric algorithm, by means of which two keys are generated—the public and private one. Users are in possession of the private one and use it to decrypt information, thus ensuring that they alone are able to obtain the data in question [18].

Hingwe and Bhanu (2014) explain a database model on the cloud with architecture consisting of two additional layers, which are used depending on whether the data is sensitive or otherwise. Data encryption is added to the layers, and this takes the form of double encryption if the information is deemed sensitive. A key is needed for this encryption and a symmetric key provided by the database server is used for such purpose. Where sensitive information is concerned and hence two layers are used, this is split into two by an algorithm to improve it [37].

Cheng et al. (2014) provide a summary of the most direct threats to which a customer of a cloud supplier may be exposed. Among these threats, the suppliers themselves may use their data for their own interests or cybercriminals may acquire the data, and they propose a simple scheme to deal with these threats involving splitting the data into fragments. After performing a hash function, these fragments are then packaged—what makes this scheme work is that these packages are randomly distributed among different storage points in such a way that they possess no useful information on their own [16].

Thilakanathan et al. (2014) provide us with a security model to be used in monitoring patients via remote devices such as mobile phones and bracelets. This model makes use of double encryption, symmetric encryption, and encryption using the ElGamal algorithm, whereby mobile devices generate the patient's data and this data is encrypted using a symmetric key. The second encryption, which is asymmetric, is used to improve security, and its function will be to encrypt the public key being used.

The disadvantage of using a symmetric key, however, is that it loses the identity of the data [39]. Subashini and Kavitha (2011) explain a security model that does not prevent the database from being hacked but rather ensures the data obtained is of no value.

They cite the example of a user's login and password in which two pieces of unrelated, separate data are of no value. Their model involves splitting the data stored into a Public Data Segment (PDS) and a Sensitive Data Segment (SDS), and SDS data needs to be fragmented still further, until each fragment does not have any value individually. This data is split using the algorithm they describe and explain in such a way that the former ceases to be of any value individually. This model is mainly focused on providing security in avoiding intrusion [38, 40, 41].

In summary, in Table 1 a comparison of the above literature based on the most important parameters as security mechanism and problem tackled is shown.

3.2. Proposed Security Solution. The literature review, the comparison, and the analysis help in proposing the theoretical approach to put into practice. The environment is explained in [42] where the proposed approach is applicable in healthcare which is the most important sector of every country around the globe. A set of information obtained from different hospitals and clinics located in the province

Table 1: Comparison of different security mechanisms in the literature.

Publication	Security mechanism	Problem tackled
Wang et al. (2009) [38]	CP-ABE scheme	Encryptor leakage (leakage of randomness)
Cho et al. (2016) [31]	Double layered architecture	Privacy invasion of personal users in big data
Liu et al. (2015) [32]	Authenticator-based data integrity verification techniques	Verification of data integrity
Fabiano et al. (2015) [33]	OpenSSL encryption package	Level of security of data
Yan et al. (2016) [34]	Proxy-based additive homomorphism	Information security of trusted suppliers
Zhou et al. (2015) [35]	Chaos stochastic process and the Line Map principle	Image security
Hsu et al. [36]	Variant of the Diffie Hellman algorithm	Secure transfer of data
Hingwe and Bhanu (2014) [37]	Data encryption or double encryption for sensitive data	Security on cloud
Thilakanathan et al. (2014) [39]	ElGamal algorithm	Security in monitoring patients via remote devices
Cheng et al. (2014) [16]	Splitting data in fragments and hash function	Security for a cloud supplier
Jing (2014) [18]	AES encryption algorithm and RSA algorithm	Security on the cloud
Subashini and Kavitha (2012) [40]	Sensitive Data Segment	Ensuring the data

of Valladolid (Spain) is proposed theoretically in this model, with information about patients being stored by them on a cloud-based storage server. Moreover, this scheme may be used to scale it to any group of hospitals or clinics in any country. To ensure security of the available systems, such as identification cards or firewalls, it is necessary to increase the security of the information being extracted and stored by these systems, because information security is extremely an important issue.

The possible threats considered are as follows: (a) the internal agent belongs to the health system and that is authorized to access the information but uses it nonethically; (b) second one is the intermediary agent, being identified as a member of the group in possession of the stored information from storage systems; and (c) the external agent is the third kind of threat, which may be anyone other than the authorized user of healthcare system. The proposed improvement focuses on improving the security against these three threats.

The information that requires protection is maintained as the set of data expressed in the form of text or images. The first theoretical proposal involves using the double-layer scheme suggested by Subashini and Kavitha (2012), in which the text data is split in such a manner that it is of no use and is of no value. By doing this we obtain a certain anonymity regarding the data, which is an important factor in the area of health [40]. As an alternative, the scheme by Cho et al. (2016) can be used, which involves splitting the data into sensitive and nonsensitive but for the same purpose, that is, to make the information anonymous. The next step will be to use the proposal made by Jing (2014), which describes the use of double encryption in order to protect data [18]. Now that the data has been split, encryption will help to ensure that intermediary and external agents will not obtain information in the format of flat text. The encryption mechanism is utilized specifically to make unauthorized access of the data impossible [43]. Additionally, the algorithm developed by Zhou et al. (2015) can be used for the medical images, due to the fact that it often tends to be preferable not to split into fragments, as these medical images are deemed to be as valuable as written reports and hence need to be protected. The proposed efficient security mechanism ensures that only authorized persons are able to decrypt the images.

In Figure 2, we visually explain the two alternatives put forward in the first stage when splitting the text-based information generated. Whereas encrypting the text and images is the same in both situations, these two alternatives are split into 3 layers, with layer 1 being the lower one that includes the text and images and in layer 3 the data is encrypted. The first layer simply represents the type of information gathered, and in this case the data is the composition of text and images.

On the left side, the technique by Cho et al. (2016) is utilized that splits the text into sensitive and nonsensitive information that represents personal details and in general labels such as "name," respectively. On the other right side, the technique given by Subashini and Kavitha (2012) is used which fragments the text when it is required and stops until it gives value [40]. Furthermore, the proposed scheme fetches the fragmented text and delivers the actual data to the authorized users and hence ensures the security.

In both cases, we maintain the images without either varying or fragmenting owing to the fact that we have at our disposal the algorithm developed by Cho et al. (2016) in order to ensure their efficient encryption and that no fragment of the image can be obtained without the suitable key [31]. All this information that has already been encrypted will be passed via Internet to the cloud servers.

4. Conclusion

This article provides different mechanisms and algorithms used to ensure big data security. Some of these techniques to help preserve information security are data modification techniques, cryptographic methods, protocols for data

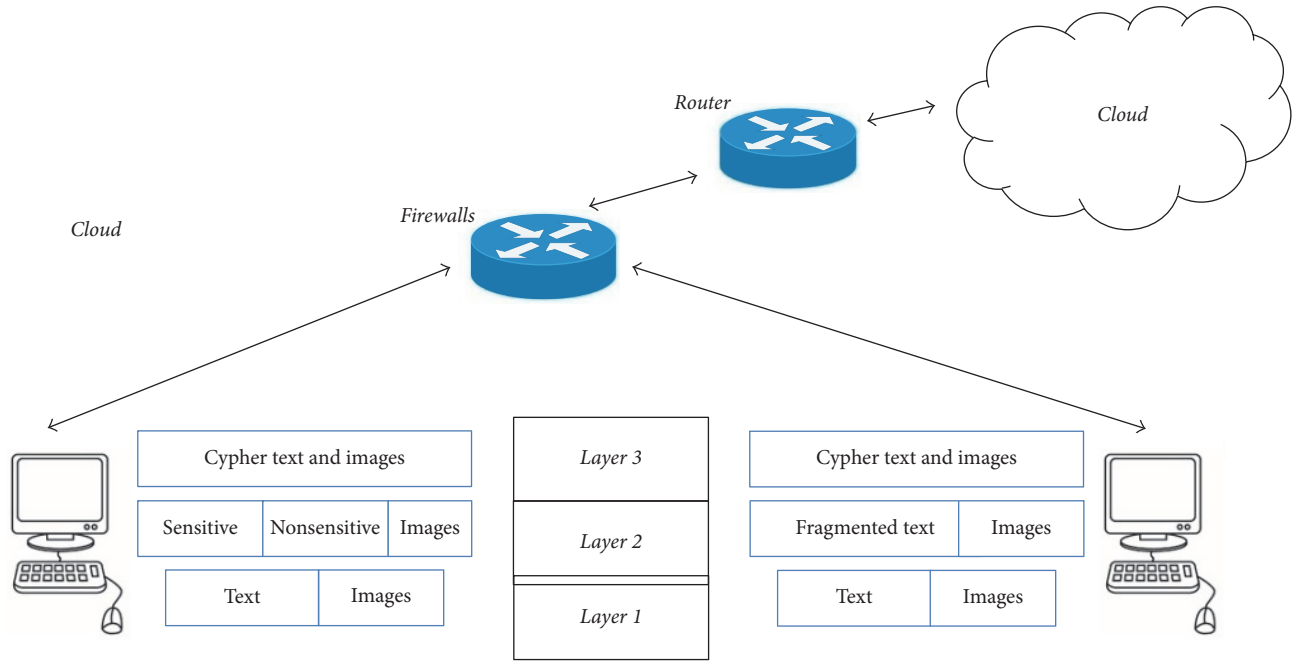


Figure 2: Proposed security scheme.

sharing, and query auditing methods. Although there remains much to do in the field of big data security, research in this area is moving forward, from both a scientific and commercial point of view. Two security alternatives have been proposed theoretically. The new proposed model pretends to give more security on the cloud in hospitals and clinics in Valladolid, Spain.

There is no perfect security system, as the methods currently in use are meant for other applications. Technology has been taking huge steps forward over the years, which may help to create algorithms that cannot currently be used owing to the computational load they require. However, this technology is the same for hackers, meaning that they need increasingly less time to discover the keys. Hence, is security perfect? At present, the only way of remaining beyond the reach of cybercriminals is not to be on the Net, although this of course is not a solution. This is because we are talking about storing data on the cloud, which is something that can be accessed via the Net and whereby disconnecting would mean not gaining access to that data or simply not being able to store it. Should we combine all these mechanisms and security algorithms? This might be one solution; it seem that things depend to quite a large extent on situations in so far as some are health-oriented; others are geared to protecting a database and others to protecting the very keys that encrypt the data being stored. One of these ways is the one we have proposed here which, although far from being perfect, may help to prevent cyber-attacks from the man-in-the-middle.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research has been partially supported by the European Commission and the Ministry of Industry, Energy and Tourism under the project AAL-20125036 named "Wetake-Care: ICT-Based Solution for (Self-) Management of Daily Living," by National Funding from the Fundação para a Ciência e a Tecnologia (FCT) through the UID/EEA/500008/2013 Project, by the Government of the Russian Federation, Grant 074-U01, and by Finep, with resources from Funttel, Grant no. 01.14.0231.00, under the Centro de Referência em Radiocomunicações (CRR) project of the Instituto Nacional de Telecomunicações (Inatel), Brazil.

References

[1] JM. Martínez Sesmero, ""Big Data"; application and utility for the healthcare system," *Farm Hosp*, vol. 39, no. 2, pp. 69-70, 2015.

[2] D. Shin, T. Sahama, and R. Gajanayake, "Secured e-health data retrieval in DaaS and Big Data," in *Proceedings of the 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services, (Healthcom '13)*, pp. 255–259, IEEE, Lisbon, Portugal, October 2013.

[3] V. A. Chang, "A model to compare cloud and non-cloud storage of Big Data," *Future Generation Computer Systems*, vol. 57, pp. 56–76, 2016.

[4] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, "Promises and challenges of big data computing in health sciences," *Big Data Research*, vol. 2, no. 1, pp. 2–11, 2015.

[5] C. L. P. Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[6] D. Agrawal, A. El Abbadi, V. Arora et al., "Mind your Ps and Vs: a perspective on the challenges of big data management

and privacy concerns," in *Proceedings of the 2015 International Conference on Big Data and Smart Computing, (BIGCOMP '15)*, pp. 1–6, Republic of Korea, February 2015.

[7] B. Logica and R. Magdalena, "Using big data in the academic environment," *Procedia Economics and Finance*, vol. 33, pp. 277–286, 2015.

[8] A. Gandomi and M. Haider, "Beyond the hype: big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144, 2015.

[9] X. Jina, B. Waha, X. Chenga, and Y. Wanga, "Significance and challenges of big data research," *Big Data Research*, vol. 2, pp. 59–64, 2015.

[10] P. Sommer, "DI commentary: big data and privacy," *Digital Investigation*, vol. 15, pp. 101–103, 2015.

[11] Z. Goli-Malekabadi, M. Sargolzaei-Javan, and M. K. Akbari, "An effective model for store and retrieve big health data in cloud computing," *Computer Methods and Programs in Biomedicine*, vol. 132, pp. 75–82, 2016.

[12] MongoDB, https://www.mongodb.com.

[13] Cassandra Apache, "Apache software foundation," http://cassandra.apache.org.

[14] Google BigTable, "Google cloud platform," https://cloud.google.com/bigtable.

[15] W. Tian and Y. Zhao, "Big data technologies and cloud computing," *Optimized Cloud Resource Management and Scheduling Theory and Practice*, pp. 17–49, 2015.

[16] H. Cheng, W. Wang, and C. Rong, "Privacy protection beyond encryption for cloud big data," in *Proceedings of the 2nd International Conference on Information Technology and Electronic Commerce, (ICITEC '14)*, pp. 188–191, IEEE, Dalian, China, December 2014.

[17] F. F. Moghaddam, M. B. Rohani, M. Ahmadi, T. Khodadadi, and K. Madadipouya, "Cloud computing: vision, architecture and characteristics," in *Proceedings of the 6th IEEE Control and System Graduate Research Colloquium, (ICSGRC '15)*, pp. 1–6, IEEE, Shah Alam, Malaysia, August 2015.

[18] P. Jing, "A new model of data protection on cloud storage," *Journal of Networks*, vol. 9, no. 3, pp. 666–671, 2014.

[19] N. Kumar, A. V. Vasilakos, and J. J. Rodrigues, "A multi-tenant cloud-based DC nano grid for self-sustained smart buildings in smart cities," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 14–21, 2017.

[20] I. de la Torre-Díez, B. Garcia-Zapirain, M. López-Coronado, and J. J. Rodrigues, "Proposing telecardiology services on cloud for different medical institutions: a model of reference," *Telemedicine and e-Health*, vol. 23, no. 8, pp. 654–661, 2017.

[21] Y. Wen, X. Zhu, J. J. P. C. Rodrigues, and C. W. Chen, "Cloud mobile media: reflections and outlook," *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 885–902, 2014.

[22] C.-W. Tsai and J. J. P. C. Rodrigues, "Metaheuristic scheduling for cloud: a survey," *IEEE Systems Journal*, vol. 8, no. 1, pp. 279–291, 2014.

[23] R. Samani, B. Honan, and J. Reavis, "Cloud security alliance research," *CSA Guide to Cloud Computing*, pp. 149–169.

[24] D. Puthal, S. Nepal, R. Ranjan, and J. Chen, "A dynamic prime number based efficient security mechanism for big sensing data streams," *Journal of Computer and System Sciences*, vol. 83, no. 1, pp. 22–42, 2017.

[25] M. R. Aswin and M. Kavitha, "Cloud intelligent track - Risk analysis and privacy data management in the cloud computing," in *Proceedings of the International Conference on Recent Trends in Information Technology, (ICRTIT '12)*, pp. 222–227, IEEE, Chennai, India, April 2012.

[26] D. Hodges and S. Creese, "Breaking the arc: risk control for big data," in *Proceedings of the 2013 IEEE International Conference on Big Data, (Big Data '13)*, pp. 613–621, IEEE, Santa Clara, California, USA, October 2013.

[27] IEEE Xplore Digital Library, http://ieeexplore.ieee.org/search/advsearch.jsp.

[28] Scopus, http://www.scopus.com.

[29] Science Direct, http://www.sciencedirect.com.

[30] Web of Science, https://apps.webofknowledge.com/UA_General-Search_input.do?product=UA&search_mode=GeneralSearch&SID=X1MambjOfgrnGY3sm74&pref.

[31] D.-E. Cho, S. J. Kim, and S.-S. Yeo, "Double privacy layer architecture for big data framework," *International Journal of Software Engineering & Applications*, vol. 10, no. 2, pp. 271–278, 2016.

[32] C. Liu, C. Yang, X. Zhang, and J. Chen, "External integrity verification for outsourced big data in cloud and IoT: a big picture," *Future Generation Computer Systems*, vol. 49, pp. 58–67, 2015.

[33] E. Fabiano, M. Seo, X. Wu, and C. C. Douglas, "OpenDBDDAS toolkit: secure mapreduce and hadoop-like systems," *Procedia Computer Science*, vol. 51, pp. 1675–1684, 2015.

[34] Z. Yan, W. Ding, V. Niemi, and A. V. Vasilakos, "Two schemes of privacy-preserving trust evaluation," *Future Generation Computer Systems*, vol. 62, pp. 175–189, 2016.

[35] G. Zhou, D. Zhang, Y. Liu, Y. Yuan, and Q. Liu, "A novel image encryption algorithm based on chaos and line map," *Neurocomputing*, vol. 169, pp. 150–157, 2015.

[36] C. Hsu, B. Zeng, and M. Zhang, "A novel group key transfer for big data security," *Applied Mathematics and Computation*, vol. 249, pp. 436–443, 2014.

[37] K. K. Hingwe and S. M. S. Bhanu, "Sensitive data protection of DBaaS using OPE and FPE," in *Proceedings of the 4th International Conference on Emerging Applications of Information Technology, (EAIT '14)*, pp. 320–327, Kolkata, India, December 2014.

[38] C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring data storage security in cloud computing," in *Proceedings of the 17th International Workshop on Quality of Service (IWQoS '09)*, pp. 1–9, IEEE, Charleston, SC, USA, July 2009.

[39] D. Thilakanathan, Y. Zhao, S. Chen, S. Nepal, R. A. Calvo, and A. Pardo, "Protecting and Analysing Health Care Data on Cloud," in *Proceedings of the 2nd International Conference on Advanced Cloud and Big Data, (CBD '14)*, pp. 143–149, IEEE, Huangshan, China, November 2014.

[40] S. Subashini and V. Kavitha, "A metadata based storage model for securing data in cloud environment," *American Journal of Applied Sciences*, vol. 9, no. 9, pp. 1407–1414, 2012.

[41] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," *Journal of Network and Computer Applications*, vol. 34, no. 1, pp. 1–11, 2011.

[42] I. De La Torre-Díez, M. Lopez-Coronado, B. Garcia-Zapirain Soto, and A. Mendez-Zorrilla, "Secure cloud-based solutions for different eHealth services in spanish rural health centers," *Journal of Medical Internet Research*, vol. 17, no. 7, article no. e157, 2015.

[43] Z. Wang, C. Cao, N. Yang, and V. Chang, "ABE with improved auxiliary input for big data security," *Journal of Computer and System Sciences*, vol. 89, pp. 41–50, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2017, Article ID 6274824, 8 pages
https://doi.org/10.1155/2017/6274824

Research Article

The Fusion Model of Multidomain Context Information for the Internet of Things

Bing Jia,[1] Shuai Liu,[1] Yushuai Guan,[1] Wuyungerile Li,[1] and Weiwu Ren[2]

[1]*College of Computer Science, Inner Mongolia University, Hohhot, China*
[2]*School of Computer Science and Technology, Changchun University of Science and Technology, Changchun, China*

Correspondence should be addressed to Shuai Liu; cs_liushuai@imu.edu.cn

Received 20 August 2017; Accepted 11 October 2017; Published 13 November 2017

Academic Editor: Yin Zhang

The Internet of Things aims to provide the user with deep adaptive intelligence services according to the user's personalized characteristics. Most of the characteristics are presented in the form of high-level context. But it often lacks methods to obtain high-level context information directly in the Internet of Things. In this paper, so as to achieve the corresponding high-level context information using the specific low-level multidomain context directly obtained by different sensors in the Internet of Things, we present a machine learning method to construct a context fusion model based on the feature selection algorithm and the multiclassification algorithm. First, we propose a wrapper feature selection method based on the genetic algorithm to obtain a simpler and more important subset of the context features from the low-level multidomain context, by defining a suitable fitness function and a convergence condition. Then, we use the decision tree algorithm which is a multiclassification algorithm, based on the rules obtained by training the subset of context features, to determine which high-level context the record set of the low-level context information belongs to. Experiments confirm that the model can be used to achieve higher classification accuracy without more significant time consumption.

1. Introduction

The Internet of Things technology is a network expanded to Internet-enabled objects, whose main function is to connect these objects [1]. It not only greatly improves the convenience of networks but also meets needs that people could not imagine before [2, 3]. The Internet of Things is full of various kinds of information about communication, sensing, and computing information to provide the user with more intelligence services [4]. The context data which is produced by the process of providing intelligence services by sensors and intelligent devices in the Internet of Things is massive and valuable, because the context data can be used to affect the human's service experience in many different ways. The "context" is not a new concept, but up till now, no unified definition of "context" exists either theoretically or practically. This is because different researchers put forward different definitions based on different backgrounds, different understandings, and different perspectives [5]. Schilit and Theimer proposed the concept of "context" [6] first, and they thought that the context information includes the user's position, the user's identity, the physical objects around the user, and the interaction state of the devices used by the user. Then, many researchers [5–9] proposed the definition of context information based on their own research field and perspectives. On the whole, these definitions were defined based on the traditional viewpoint of "user-center," which mainly consists of three basic essential elements: human, machine, and environment. However, for the Internet of Things, the context information should be coordinated between two humans, a human and an object, and two objects. We adopt the definition in reference documentation [10], which regards context information as the interaction information between human, object, machine, and environment in the Internet of Things. It contains both the preset static information and the dynamic information caused by the interaction. In order to adapt to the user's personalized needs, the provided service should have significant and personalized characteristics to

adapt to the user's context feature [11]. Usually, during the lifetime of the service in the Internet of Things, when the user changes the low-level multidomain context, such as location, temperature, or illumination, the high-level context for the user's personalized characteristics may always be changing [12]. So, it is vital to get the high-level context feature by fusing the low-level context information of the user's multidomain environment timely [13].

Context fusion is a process to obtain the high-level context by dealing with the multidomain low-level context (which consists of monitoring, sensing data, and so on) based on some methods and prior knowledge. Nowadays, most researches have adopted the method of rule reasoning. For example, the middleware of Context Toolkit [14] supported context reuse and customization by the abstract representation. Through encoding program for the logical rules, Context Toolkit achieved conflict detection and reasoning by OOPS (Organized Option Pruning System). Gaia [15] adopted CORBA (Common Object Request Broker Architecture), which is a distributed component architecture based on the thought of operating system, to achieve efficient reasoning by using a first-order predicate logic model. CORBA also introduced the idea of probability and fuzzy logic to deal with the uncertainty. The project of PACE (Pervasive, Autonomic Context-aware Environments) introduced three-valued logic to handle some uncertainty information based on the graphical context modeling language CML (Context Modeling Language) [16]. The SOCAM (Service-Oriented Context-Aware Middleware) [17] based on OSGI (Open Service Gateway Initiative) and the context application middleware CoBrA (Context Broker Architecture) based on agent [18] have both used OWL ontology to achieve reasoning. The expression of the rule-based reasoning is direct, unified, and accurate. This method is more suitable for solving small-scale datasets, but it is difficult for it to deal with complex systems and large-scale datasets, because the relationships between the observed symptoms and the corresponding diagnoses in large-scale systems are more complex, so it is difficult to sum up the effective rules in view of the experience of experts. This paper has introduced a method of machine learning to construct a context fusion model in order to realize the fusion processing of large-scale and multidomain context information.

2. The Classification of the Context

There is no unified standard for the classification of the context. Different researchers proposed different partition methods based on their own research backgrounds, applications, and research requirements. Dey et al. proposed that the context information was location, identity, activity, and time. Context was defined as user and role, process and mission, position, time, and equipment by Kaltz et al. [19], in order to make an extensive range of the mobile and network context [20]. Schmidt et al. proposed that the context information was the human factor and the physical environment in [21]. Luo et al. [22] made a more detailed distinction about source, purpose, and varying frequency of the context information. We analyzed the context information which may be produced in the new networking environment and the processes of service registration, service discovery, service selection, and service composition based on the needs of the user under the Internet of Things and defined the context information as the environment context information, the device context information, the user context information, and the calculation context information. Details are as follows.

2.1. The Environment Context Information. It is used to describe the context of service environment in the Internet of Things [23], including the specific environment, the scene environment, and the temporal environment. The specific environment refers to the measurable environmental status (such as temperature, humidity, and sound and light noisiness). The scene environment refers to a relatively stable environment (such as conference room or a café). The temporal environment includes some temporal objects (such as time point and time period) and the relations among temporal objects [24].

2.2. The Device Context Information. It mainly refers to the ability and outline of a device in this paper. The device is the carrier of the interaction between the service and the user, including the static device context and the dynamic device context. The static device context includes the device type context and the display performance. The dynamic device context includes the signal intensity, the moving speed, and the electricity of the device.

2.3. The User Context Information. It mainly refers to the ability and outline of the user in this paper. The user refers to the entity which can initiate the service demand in the Internet of Things (people, an ordinary object, etc.). It includes the static user context, the dynamic user context, and the historical user context. The static user context includes the user's identification, identity, and preferences. The dynamic user context includes the user's position, posture, and permission. The position includes the geometric position (specific latitude, longitude, and altitude obtained through GPS, etc.) and the relative position ("end of corridor," "the east side of the classroom," etc.).

2.4. The Calculation Context Information. It mainly refers to the feature associated with the calculation property, including the static calculation context and the dynamic calculation context. The static calculation context includes the computing capacity (storage capacity, processor capacity, etc.) and the network capacity (network type, network bandwidth, communication cost, etc.). The dynamic calculation context includes the CPU utilization rate, the memory utilization rate, and the communicable neighbor nodes.

3. Context Fusion Model

3.1. Structure of the Context Fusion Model. Some of the various contexts mentioned above can be directly and originally

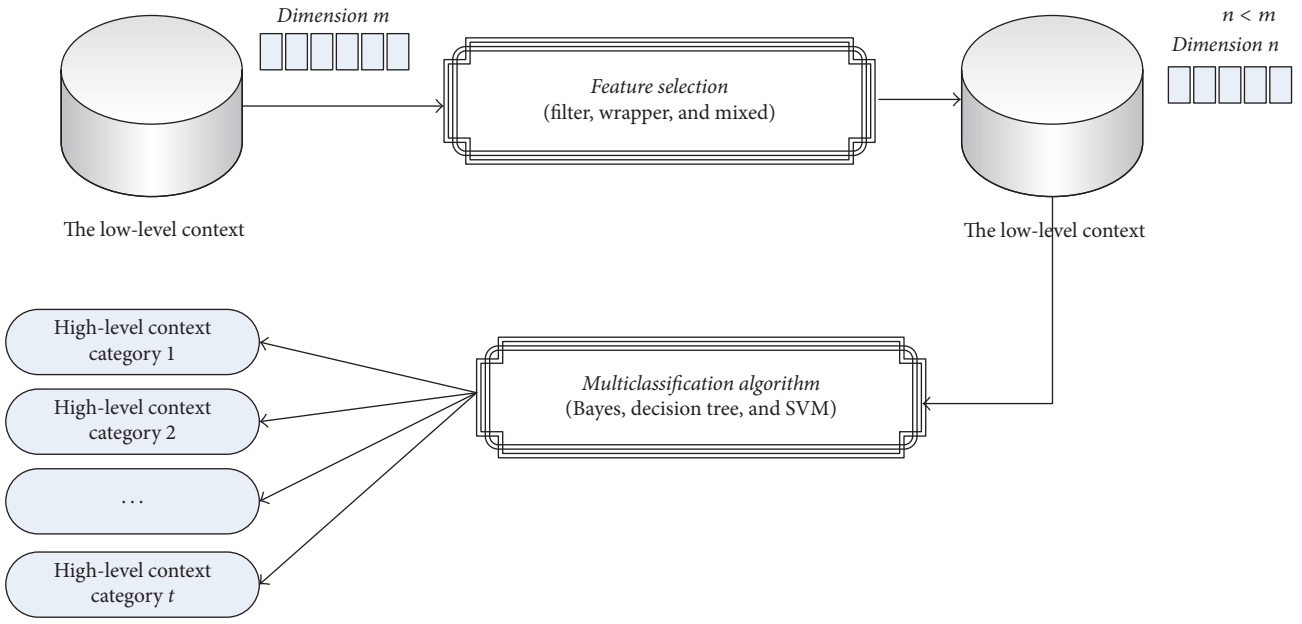


Figure 1: The structure context fusion model.

obtained (e.g., light, remaining battery), which are called the low-level contexts.

The original acquisition is the process of obtaining various types of context information (or data) directly from a variety of context information sources (such as sensors, RFID readers, and cameras). The others need to integrate the low-level context information which was obtained from the information source to obtain the high-level context information (scenes, relative position, etc.). For example, a scene which the user is faced with cannot be obtained directly, and it must be a comprehensive analysis of the low-level context, such as the temperature, the humidity, the speed, the position, and the direction, to infer which scene the user faced (in the corridor, upstairs, etc.). The context information of the perception is often complex and has a multidomain for the Internet of Things, and it is difficult to sum up the effective rules based on the experience of the experts. So, it is not suitable to use the rule-based method for reasoning. This paper has introduced the machine learning method and given the context fusion model based on the genetic algorithm and the decision tree. The model structure is shown in Figure 1. First, the feature selection method is adopted to deal with the input dimension reduction. Then, it can determine the kind of high-level context by fusing the low-level context based on the classification algorithm. This method can be trained to obtain inference rules through a large number of samples, and it does not require human intervention. Although the inference result is not very accurate, it is easy to implement and apply in processing large-scale information. It will be more feasible if the requirements of identification accuracy are not very strict.

The model contains two parts: feature selection and classification. The purpose of the feature selection is to decrease the dimension of the training samples and delete some features which are unrelated or weakly related to the task, in order to obtain the simple but important feature subsets [25]. The sample can be trained to get classification rules on the basis of the feature subsets. Then, according to the classification rules, some low-level contexts can be classified as a kind of high-level context. In this model, dozens of dimensions (recorded as "m") of the low-level context information constitute the original input space, and the high-level context corresponding to the low-level context information (recorded as "t") constitutes the output space. We can obtain n ($n < m$)-dimensional compact feature subset by the feature selection algorithm (filter, wrapper, or mixed mode) for the original input space. Then, we can reduce the dimensions of the test data based on the feature subset. Finally, the classification and recognition task of each sample is completed by the multiclassification algorithm (Bayes, decision tree, or SVM).

Bayes classification follows Bayes theorem. Bayes theorem gives a method to calculate the posterior probability. Bayes classification provides a method which can combine practical learning algorithms and prior knowledge and observed data. It provides a beneficial perspective for understanding and evaluating many learning algorithms. The naive Bayes classification is the most commonly used method in Bayes classification. As the name suggests, this classifier uses the naive Bayes theorem to get the classification for a given variable value. The naive Bayes classifier is a very simple classification algorithm based on probability models with independence assumptions between predictors. The independence assumptions do not often have an impact on reality. Therefore, they are considered as naive. A naive Bayes model is very useful for large datasets, which is easy to build and with no complicated iterative parameter. Despite its simplicity, the naive Bayes classifier is widely used, because it usually behaves well and often outperforms more sophisticated classification methods.

Support vector machine (SVM) is a set of supervised learning methods used for classification, regression analysis, and outliers detection, which is derived from the statistical learning theory. It often yields great classification results from complex and noisy data. SVM is mainly used for two categories of classification problems. Although some of the papers mentioned that the K support vector machine combination can be used to solve the K class classification problem, this process requires some caution.

In the classification technique, the decision tree is a powerful classification method. It can be used to determine the characteristics of the training data segmentation, resulting in a good generalization. Decision tree algorithm can naturally deal with binary or multiclassification problems. And the leaf nodes can refer to any of the K classes concerned.

Through a lot of experiments, we find that the decision tree classification algorithm and the wrapper feature selection method based on the genetic algorithm have achieved good classification results in the context information for the Internet of Things. In this paper, the experiment will be given later in Section 4. Next, we will introduce the feature selection method based on the genetic algorithm and the multiclassification method of the decision tree.

3.2. Feature Selection Based on the Genetic Algorithm. Generally, there are three modes of feature selection, namely, the filter mode, the wrapper mode, and the mixed mode. The filter mode uses the properties of the data itself as an evaluation index for feature subset, and the wrapper uses the correct rate of the machine learning algorithm as the evaluation index of the feature subset [26]. In general, the filter feature selection is faster. The process of the selection is not related to the machine learning algorithm, so the feature subset may not be adaptable to the certain machine learning algorithm. This makes the result subset after feature selection not necessarily the optimum one. The wrapper mode is slower than the filter, because it needs to do cross-certification and more complex calculation. The feature subset can be adapted to the classification algorithm, so the selection result is generally better. The mixed mode needs to do feature selection in two steps, so the computation time is very large while the accuracy rate is not improved significantly. So, the mixed mode is not used commonly. This paper has chosen the wrapper model to do the feature selection. The working principle of the wrapper is that it needs to package data into different feature subsets in accordance with dimensionality and make its selection through the correct classification rate. So, we need to search the whole feature subset space.

Search strategies are generally divided into three types, namely, exhaustive search, heuristic search, and uncertain search [27]. The exhaustive search strategy can search for all possible feature subsets, and it will be able to find the optimal subset of features. But it is difficult to achieve the optimal solution for a large number of features, because the cost of the space and time is large. The heuristic search strategy will search for subset features according to a certain heuristic rule [28]. Its cost is less, but it is liable to fall into a local optimum, and the global optimum cannot be obtained. The uncertainty search strategy is a balance of the above two kinds of search, for example, the genetic algorithm. We use the genetic algorithm in this paper. Figure 2 shows the wrapper mode based on the genetic algorithm in the feature selection process. Because the context classification is a kind of multiclassification, the typical methods of Bayes, decision tree, and SVM are chosen as the evaluation function of feature subset. The decision tree is the main method used in this paper; Bayes and SVM are the methods used for comparative experiments.

In Figure 2, the parameter needs to be initialized first, which is a key of the genetic algorithm. The feature subset is coded as the only parameter in this paper. Then, the model completes the generation of feature subset, the evaluation of feature subset, evaluation stopping, and the verification of the result. An initial subset is randomly generated according to the initial parameters and the original population. The characteristic subsets of each generation are generated according to the relevant parameters calculated by the genetic operator. The fitness function is defined as follows:

$$\text{fit}(\alpha_i) \begin{cases} \text{CR}(\alpha_i) - \dfrac{\text{CR}_{\min}}{|S|} & \text{if } \text{CR}_{\max} + \text{CR}_{\min} \le 2 * \text{CR}_{\text{avg}} \\ \text{CR}(\alpha_i) + \text{CR}_{\max} - 2 * \dfrac{\text{CR}_{\text{avg}}}{|S|} & \text{if } \text{CR}_{\max} + \text{CR}_{\min} > 2 * \text{CR}_{\text{avg}}. \end{cases} \tag{1}$$

In (1), α_i is the ith generation individual. $\text{CR}(\alpha_i)$ is the correct classification rate of the subset. S is all subsets in this generation. $|S|$ is the number of subsets. $\text{CR}_{\max}$ is the maximum classification accuracy of the subset in the ith generation. $\text{CR}_{\min}$ is the minimum classification accuracy of the subset in the ith generation. CR_{avg} is the average classification accuracy of the subset in the ith generation.

Genetic operators are the key to implement the optimal search. There are three kinds of operators, namely, selection, crossover, and mutation [29]. The selection operator selects the individual which can be inherited to the next generation by a certain strategy, based on the evaluation of individual fitness [30]. The crossover operator randomly selects two chromosomes according to a certain crossover probability and exchanges some of its genes in some way to form two new individuals. The mutation operator is used to generate a new individual by randomly changing some bits on the chromosome in a small probability P_m. The crossover operation is the main method of generating new individuals in the process of genetic operation. It determines the global

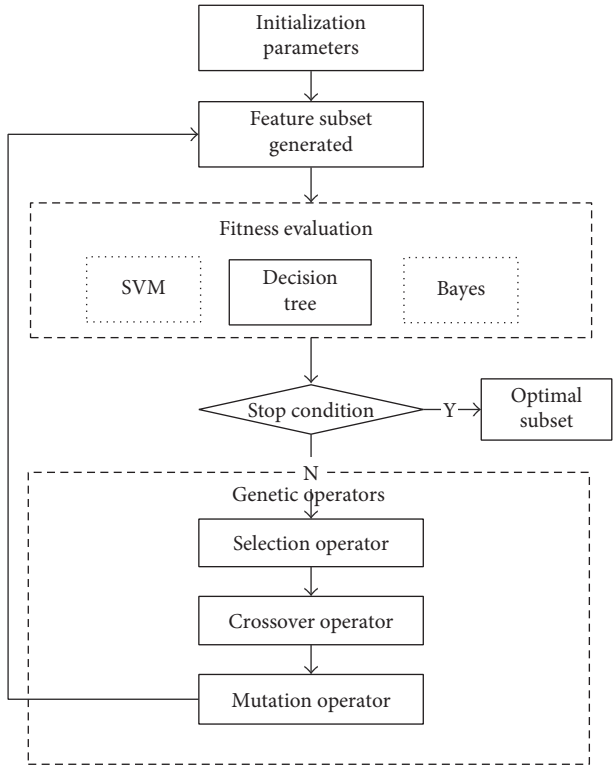


Figure 2: Feature selection process based on the genetic algorithm.

search capability [31] of the genetic algorithm. The mutation operation is an auxiliary method to generate new individuals, and it determines the local search ability of the genetic algorithm. The mutual cooperation between the crossover operator and the mutation operator will complete the global search and local search for the search space. It can make the genetic algorithm complete the search process with a good performance [32]. In this paper, two conditions of the algorithm convergence are designed: one is that the subset has achieved stability and the other is that generation quantity has been over the threshold.

3.3. Decision Tree Classification Algorithm. One of the commonly used methods in data mining is the decision tree learning method. The goal of the decision tree learning method is to create a model that predicts the value of the target variable based on several input variables. Each internal node corresponds to an input variable and there are edges to children for each possible input variable value. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

The top node of the tree is root. The decision tree can be established and redefined by the training of test samples. The building process of the decision tree is a machine learning process [33].

Currently, there are lots of classic decision tree algorithms, such as the ID3 algorithm, the C4.5 algorithm, and the CART algorithm. The ID3 algorithm can only deal with discrete data. The C4.5 algorithm has made some improvements to the ID3 algorithm according to the information gain ratio which is used to select the test attributes. Some of these are as follows: dealing with both continuous and discrete attributes, dealing with training data with missing attribute values, dealing with attributes with differing costs, and pruning trees after creation [33]. The CART algorithm cannot efficiently handle large-scale training sample data. Based on the above analysis, in this paper, we have chosen the C4.5 algorithm.

3.4. C4.5 Algorithm

3.4.1. The Calculation of the Information Gain. If the value of the property (recorded as "a") of the sample will divide the sample set (recorded as "T") into m subsets, namely, $T_1, T_2, \ldots, T_m$, the formula for calculating the information gain is as follows:

$$\text{gain}(a) = \text{info}(T) - \sum_{i=1}^{m} \frac{|T_i|}{|T|} \times \text{info}(T_i). \tag{2}$$

In (2), "$|T|$" is the number of samples in dataset "T"; "$|T_i|$" is the number of samples in subset "T_i"; info(T) is calculated as follows:

$$\text{info}(T) = -\sum_{j=1}^{s} \text{freq}(C_j, T) \times \log_2 \left(\text{freq}(C_j, T)\right), \tag{3}$$

where freq(C_j, T) is the frequency of the category of the sample data "C_j" and "s" is the number of categories of the sample "T."

3.4.2. The Calculation of the Information Gain Ratio. One has

$$\text{gainratio}(a) = \frac{\text{gain}(a)}{\text{splitinf}(a)}. \tag{4}$$

In (4), splitinf(a) represents the split information, which is the potential information generated when "T" is divided into "h" parts; the formula is as follows:

$$\text{splitinf}(a) = -\sum_{i=1}^{h} \frac{|T_i|}{|T|} \times \log_2 \left(\frac{|T_i|}{|T|}\right). \tag{5}$$

3.5. Building the Decision Tree. The method of building a decision tree is proposed as follows:

S1: Create the node "N" and start building the decision tree from the node.

S2: If the samples are in the same class, the node becomes a leaf node. Label the node with this class.

S3: Otherwise, for each property, the data should be dispersed if its data is continuous.

S4: The information gain ratio is calculated for each attribute, and then the property which has the highest information gain ratio will be selected and labeled.

S5: The consistent value is calculated for the properties of each branch. And then it produces the branch with the same value.

S6: Let "S" be the branch set of the training test set. If "S" is empty, it needs to add a leaf node and be marked by the class.

S7: If "S" is not empty, go to "S4."

To prevent overfitting between the established trees and the training samples to enhance the speed and accuracy of the subsequent classification, we usually need a pruning strategy. Pruning is a technique in machine learning. It reduces the scale of decision trees by eliminating sections of the tree that offer small power to classify instances. Pruning not only reduces the complexity of the final classifier but also improves the predictive accuracy by the reduction of overfitting. Common methods of calculating the classification error rate and the encoding length of the decision tree are used to prune the decision tree [34]. For each nonleaf node, the pruning method will calculate the expected classification error rate if the node is pruned based on the classification error rate. At the same time, according to the classification error rate of each branch and the weight of each branch, the expected classification error rate will be calculated if the node is not pruned. If the expected error rate gets larger because of the pruning, the pruning will be abandoned, and each branch of the corresponding node will be retained. Otherwise, each branch of the corresponding node will be pruned. In the pruning process, an independent test dataset is needed to be used to evaluate the accuracy of classification for the pruned tree, to retain a pruning decision tree which is the minimum expected error rate after being pruned.

4. Experiments and Results Analysis

In this paper, we use a context information dataset, "Sensor Signal Dataset for Exploring Context Recognition of Mobile Devices" proposed in [35]. There are thirty-two columns about the sensor information [25]. Among them, the first column and the second column show the sequence number of the test scheme and the times of the test repetition; the third column shows the context information of the time; the fourth column to the ninth column show the context information of the device direction; the tenth and the eleventh columns show the context information of the device stability; the twelfth column shows whether the device is in the hand; the thirteenth column to the nineteenth column show the context information of the illumination; the twentieth column to the twenty-third column show the context information of the temperature; the twenty-fourth column to the twenty-sixth column show the context information of the humidity; the twenty-seventh column to twenty-ninth column show the context information of the noise; the thirtieth column to the thirty-second column show the action context information of human behavior. The 10,470 records from the third column to the thirty-second column are chosen in this experiment, to constitute the original context dataset ($10470 * 30$). According to the image materials given by the authors [28], we organize the dataset into eight different kinds of context scenes which are shown in Table 1.

Table 1: Eight kinds of context scenes.

Scene context name	Class code
The equipment is on the table	1
The equipment is in the hands of man	2
Walking in the office	3
Walking in the corridor	4
Going downstairs	5
Walking in the coffee shop	6
Walking in the streets	7
Going upstairs	8

Table 2: Result of feature selection (E1).

Algorithm	Feature subset
Bayes	3, 5, 7, 9, 10, 11, 12, 13, 24, 27, 28, 30
Decision tree	3, 4, 9, 10, 14, 15, 18, 19, 20, 21, 23, 25, 29
SVM	3, 5, 6, 7, 10, 11, 15, 16, 17, 20, 28, 30, 32

Table 3: Real-time comparison (E1).

Algorithm	Training time	Test time
Bayes	9.4 s	0.3 s
Decision tree	10.6 s	0.4 s
SVM	27.3 s	2.1 s

In the first experiment (E1), 2,960 records have been selected randomly as a feature selection dataset. The remaining data has been used in the classification experiment. The Bayes algorithm and the SVM algorithm have been used as comparison algorithms. The former uses a more representative method called naive Bayes. The latter uses the SVM multiclassification algorithm based on the voting mechanism and uses the C-SVM algorithm for each binary classification, whose kernel function is an RBF kernel function. The test results are as follows.

The results of feature subset selection of the three algorithms are shown in Table 2. From the table, the SVM algorithm and the decision tree algorithm select thirteen features, but the Bayes algorithm selects 12 features. We can see that the features in the third and tenth column are more important because these three algorithms all have selected these features.

Table 3 shows the time consumed by the three algorithms when training and testing. It can be seen that Bayes algorithm takes the least time, and the decision tree algorithm takes more time than Bayes. The SVM consumes more training and testing time than Bayes and the decision tree.

A comparison of the classification accuracy of the three algorithms is shown in Figure 3. As can be seen, the highest is the decision tree whose classification accuracy rate is more than 95%. The second is the SVM with nearly 89%. The lowest is the Bayes method with only approximately 57%.

So as to verify the adaptability of the proposed fusion model in the paper for different context data, we have carried

Table 4: Result of feature selection (E2).

Algorithm	Feature subset
Bayes	3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 18, 20, 25, 27, 29
Decision tree	2, 3, 5, 7, 10, 11, 12, 15, 17, 18, 19, 22, 25, 26, 30, 31
SVM	2, 3, 24, 28, 30

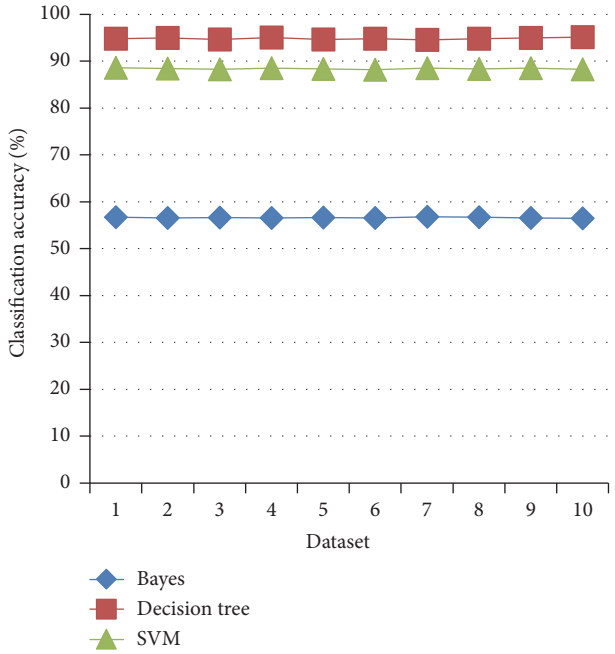


Figure 3: Comparison of the classification accuracy (E1).

out another experiment (E2) using another 980 random records in the datasets. We also used the Bayes algorithm and SVM algorithm as comparison algorithms. The results of feature subset selection of the three algorithms are shown in Table 4. The results of classification accuracy of the three algorithms are shown in Figure 4.

Simulation results show that the multiclassification algorithm based on decision tree can achieve higher classification accuracy, compared to the classical Bayes and the SVM classification algorithm. It is suitable for solving the fusion of large-scale multidomain low-level context information. So, it can obtain the high-level context class represented by the low-level context information more quickly and accurately. In addition, the random forest is an extended version of the decision tree, so if the decision tree was replaced by random forest in the context fusion model, it can also achieve very good results.

5. Conclusion

This paper has proposed a context fusion model based on the machine learning method to achieve the fusion of the multidomain low-level context information under the Internet of Things. First, the dimensions of the original data are reduced by using the wrapper feature selection method based on the genetic algorithm. Then, based on the decision

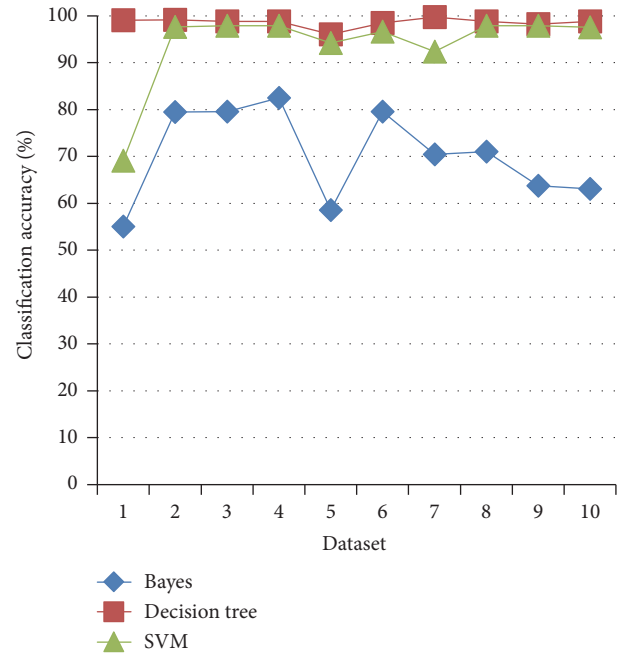


Figure 4: Comparison of the classification accuracy (E2).

tree classification algorithm, it completes the classification and recognition of the low-level context identification to determine which kind of high-level context it belongs to. The experimental results confirm the validity of the proposed model. If the recognition accuracy requirements are not particularly strict, the model will be feasible for large-scale context information. Further research is needed on the optimal selection of some related parameters in the algorithm.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (nos. 41761086, 61640013, and 61502254), the Program of Higher-Level Talents of Inner Mongolia University (no. 135138), and the Inner Mongolia Autonomous Region Science and Technology Innovation Guide Reward Fund Project under Grant no. 20121317.

References

[1] Q. Liu, Y. Ma, M. Alhussein, Y. Zhang, and L. Peng, "Green data center with IoT sensing and cloud-assisted smart temperature control system," *Computer Networks*, vol. 101, pp. 104–112, 2016.

[2] C. Zhang, Y. Yang, Z. Du, and C. Ma, "Particle swarm optimization algorithm based on ontology model to support cloud computing applications," *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 5, pp. 633–638, 2016.

[3] B. Jia, *Research on Semantic-based Service Architecture and Key Algorithms for the Internet of Things*, Jilin University, 2013.

[4] K. Kotis and A. Katasonov, "Semantic interoperability on the Web of things: The semantic smart gateway framework," in *Proceedings of the 2012 6th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2012*, pp. 630–635, July 2012.

[5] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "CA4IOT: Context awareness for Internet of Things," in *Proceedings of the 2012 IEEE International Conference on Green Computing and Communications, GreenCom 2012, 2012 IEEE International Conference on Internet of Things, iThings 2012 and 5th IEEE International Conference on Cyber, Physical and Social Computing, CPSCom 2012*, pp. 775–782, fra, November 2012.

[6] B. N. Schilit and M. M. Theimer, "Disseminating active map information to mobile hosts," *IEEE Network*, vol. 8, no. 5, pp. 22–32, 1994.

[7] J. L. Encarnação and J. M. Rabaey, *Journal of Mobile Communication*, Springer US, Boston, MA, 1996.

[8] D. Franklin and J. Flaschbart, "All gadget and no representation makes jack a dull environment," in *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments*, pp. 155–160, 1998.

[9] G. Chen and D. Kotz, "Survey of context-aware mobile computing research," Dartmouth ComPuter Seienee Teehnieal Report, 2002.

[10] P. Haitao, *Research on Feature Selection Algorithms in Machine Learning*, Shandong University, 2011.

[11] Y. Zhang, M. Chen, N. Guizani, D. Wu, and V. C. Leung, "SOVCAN: Safety-Oriented Vehicular Controller Area Network," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 94–99, 2017.

[12] W. Wei, H. Song, W. Li, P. Shen, and A. Vasilakos, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Information Sciences*, vol. 408, pp. 100–114, 2017.

[13] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[14] G. Pujolle, "An Autonomic-oriented Architecture for the Internet of Things," in *Proceedings of the IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*, pp. 163–168, Sofia, Bulgaria, October 2006.

[15] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. D. Johnson, "M2M: from mobile to embedded internet," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 36–43, 2011.

[16] N. Koshizuka and K. Sakamura, "Ubiquitous ID: standards for ubiquitous computing and the internet of things," *IEEE Pervasive Computing*, vol. 9, no. 4, pp. 98–101, 2010.

[17] Y. W. Kim, "Ubiquitous Sensor Network," ETPI. 2007.

[18] S. Fukunaga, T. Tagawa, K. Fukui, K. Tanimoto, and H. Kanno, "Development of ubiquitous sensor network," *Oki Technical Review*, vol. 71, no. 4, pp. 24–29, 2004.

[19] J. W. Kaltz, J. Ziegler, and S. Lohmann, "Context-aware web engineering: Modeling and applications," *Revue d'Intelligence Artificielle*, vol. 19, no. 3, pp. 439–458, 2005.

[20] "Context awareness," http://en.wikipedia.org/wiki/Context_awareness.

[21] A. Schmidt, M. Beigl, and H.-W. Gellersen, "There is more to context than location," *Computers & Graphics*, vol. 23, no. 6, pp. 893–901, 1999.

[22] J. Luo, X. Qing, and S. Chen, "Context-based triggered task model in pervasive computing," *Journal of Chinese Mini-Micro Computer Systems*, vol. 25, no. 8, pp. 1542–1545, 2004.

[23] Z. Wu, X. Tao, and J. Lv, "An ontology based dynamic context model," *Journal of Frontiers of Computer Science and Technology*, vol. 2, no. 4, pp. 356–367, 2008.

[24] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "Health-CPS: Healthcare cyber-physical system assisted by cloud and big data," *IEEE Systems Journal*, vol. PP, no. 99, 2015.

[25] P. Qian, S. Wang, and H. Yan, "Further feature based fuzzy classification for Context change awareness in pervasive computing," *Application Research of Computers*, vol. 27, no. 5, pp. 1648–1652, 2010.

[26] F. Alighardashi and M. A. Zare Chahooki, "The Effectiveness of the Fused Weighted Filter Feature Selection Method to Improve Software Fault Prediction," *Journal of Communications Technology, Electronics and Computer Science*, vol. 8, pp. 5–11, 2016.

[27] L. Hui and Y. Cao, "Study of heuristic search and exhaustive search in search algorithms of the structural learning," in *Proceedings of the 2010 2nd International Conference on MultiMedia and Information Technology, MMIT 2010*, pp. 169–171, April 2010.

[28] S. Liu, Z. Pan, W. Fu, and X. Cheng, "Fractal generation method based on asymptote family of generalized Mandelbrot set and its application," *Journal of Nonlinear Sciences and Applications. JNSA*, vol. 10, no. 3, pp. 1148–1161, 2017.

[29] A. H. Beg and M. Z. Islam, "Novel crossover and mutation operation in genetic algorithm for clustering," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation, CEC 2016*, pp. 2114–2121, July 2016.

[30] S. F. Chenoweth, J. Hunt, and H. D. Rundle, "Analyzing and comparing the geometry of individual fitness surfaces," in *International Conference on Consumer Electronics, 2005. ICCE. 2005 Digest of Technical Papers*, pp. 89-90, 2016.

[31] H. Liu and S. Wei, "Analysis on genetic operators," *Computer Technology and Development*, vol. 16, no. 10, pp. 80–82, 2006.

[32] B. Vázquez-Barreiros, M. Mucientes, and M. Lama, "ProDiGen: mining complete, precise and minimal structure process models with a genetic algorithm," *Information Sciences*, vol. 294, pp. 315–333, 2015.

[33] M. Somvanshi and P. Chavan, "A review of machine learning techniques using decision tree and support vector machine," in *Proceedings of the 2nd International Conference on Computing, Communication, Control and Automation, ICCUBEA 2016*, August 2017.

[34] L. Lin H, K. Chen, and H. Chiu R, "Predicting customer retention likelihood in the container shipping industry through the decision tree approach," *Journal of Marine Science Technology*, p. 25, 2017.

[35] J. Mäntyjärvi, J. Himberg, P. Kangas, U. Tuomela, and P. Huuskonen, "Sensor Signal Data Set for Exploring Context Recognition of Mobile Devices," in *Workshop "Benchmarks and a database for context recognition" in conjunction with the 2nd International Conference on Pervasive Computing (PERVASIVE 2004)*, Vienna, Austria, 2004.